Regulation and Function of the Nuclear Transport Factor KPNA7

Luke Timothy Oostdyk
Lancaster, Pennsylvania

B.S. Biochemistry, University of Delaware, 2014
M.S. Biological and Physical Sciences, University of Virginia, 2017

A Dissertation presented to the Graduate Faculty
of the University of Virginia for the Degree of
Doctor of Philosophy

Department of Biochemistry and Molecular Genetics

University of Virginia
August 2019

Abstract

Genetic material in eukaryotic cells is isolated inside the double membrane of the nucleus. To facilitate the myriad processes which occur within, cells need a mechanism to transport important nuclear proteins through the semipermeable pore complexes of the nuclear membrane. While multiple transport pathways exist, the best characterized relies on a short amino acid sequence termed a nuclear localization signal (NLS). These signals are diverse but typically consist of one or two clusters of basic amino acids. Transport of NLS containing proteins is mediated by a pair of transport receptors, Importin-α and Importin-β. Importin-α is an adapter protein which contains separate domains to recognize NLS and bind to Importin β. Facilitation of transport through the nuclear pore is mediated by interactions of Importin-β with proteins in the nuclear pore complex.

Nuclear transport is highly controlled through coordination of expression of the transport machinery and dysregulation is implicated in many disease states. One level of this control is through the Importin-α family of transport receptors. The human genome encodes seven isoforms of Import-α which have a conserved structure consisting of an N-terminal, auto-inhibitory, Importin-β-binding domain, and a C-terminal core made up of alpha-helical armadillo repeats used for binding nuclear localization signals. Despite high conservation of primary, secondary and tertiary protein structures, small variations between the isoforms of the Importin-α family result in differing affinities of each isoform for different NLS. Through differential expression of the Imp-α isoforms, the differences in NLS affinity between them functions to regulate the overall nuclear import capacity for a given cell. This mechanism of regulation is utilized to help promote various developmental programs and alteration of the expression of even a single isoform can negatively impact cellular processes and promote disease.

In this study, we have characterized the biochemical function and regulation of the most recently discovered Importin-α isoform, KPNA7. Our analysis in Chapter II has identified differential regulation of KPNA7 by the Importin-β binding (IBB) domain it contains. The canonical function of the IBB domain, which features clusters of basic amino acids, is to bind in the NLS binding groove in the body of the receptor and prevent the association of an NLS in the absence of Importin-β. Only in the presence of both

Importin-β and an NLS containing protein is there coordinated binding and formation of a heterotrimeric transport complex. In contrast to the other members of the Importin-α family, we have determined the IBB domain of KPNA7 displays a weaker auto-inhibitory function. KPNA7 binds strongly to Importin-β in the absence of an NLS and may adopt an open conformation in the nucleus. Furthermore, KPNA7 binds weakly to the Importin-α nuclear export factor CAS. We identified an Importin-β-dependent enhancement of NLS binding by KPNA7 but believe this is to be a mechanism which functions outside of relief from auto-inhibition. Together, these data suggest that, in addition to acting as a transport factor, KPNA7 may modulate the nuclear activity of NLS containing proteins.

In Chapter III of this work, we evaluated how mutations in KPNA7 identified in neurodevelopmental disease affect the NLS binding and transport capabilities of the receptor. These mutations result in amino acid substitutions proximal to the NLS binding groove of the receptor, and one, Glu344Gln, was determined to significantly reduce NLS binding and transport by KPNA7. This reduction was found to apply to both monopartite and bipartite NLS. We identified neuronal KPNA7 interacting proteins and characterized an interaction of KPNA7 with two heterogeneous nuclear ribonuclear proteins, hnRNP R and hnRNP U. The Glu344Gln substitution disrupted KPNA7 binding to each of the proteins. A functional bipartite NLS was identified in hnRNP R which is required for its nuclear localization. Binding and transport of this NLS, as well as a monopartite NLS in hnRNP U, were similarly reduced by the disease-associated mutation. We identified induction of KPNA7 expression during neurogenesis and evidence that regulation of KPNA7 expression in important for multiple developmental programs. Finally, investigation of the KPNA7 gene revealed a potential secondary effect of the disease-associated mutations via disruption of a binding site for the transcriptional insulator CTCF. Our data suggest a neuronal function for KPNA7 which is disrupted by disease-associated mutation.

In addition to our studies of KPNA7, in Chapter IV of this work, we investigated the use of the *E. coli* biotin-protein ligase BirA for identifying protein-protein interactions in cells. The commonly used BioID method utilizes a BirA mutant with an Arg118Gly amino acid substitution. We generated new BirA mutants with amino acid substitutions at

position 118 of the enzyme. We utilized a set of biochemical and cell biological methods to investigate the biotinylation characteristics of these BirA mutants, and identified BirA(Arg118Lys) as a mutant with qualities useful for proximity labeling experiments.

Overall, this work furthers the understanding of the biochemical regulation and function of KPNA7 protein and determines how mutations associated with neurodevelopmental disease affect the characteristics we have described.

Acknowledgments and Dedication

The work contained in this dissertation would not have been possible without the contributions and overwhelming support of my family, friends, colleagues, and advisors who have been instrumental in my development as both a scientist and a person over the past 5 years.

I would like to thank my friends near and far for the impact you have had on my life to keep me grounded during the grind of life as a graduate student. Thanks to the awesome soccer community in Charlottesville which has given me a place to unwind and continue to enjoy the game. Special thanks to my girlfriend Kaitlyn for her support and encouragement. I am excited to see where life takes us.

A fertile research environment is critical for any project to succeed. I'd like to thank everyone in the Center for Cell Signaling and the Department of Biochemistry and Molecular Genetics for making Virginia a great, collaborative place to do science. Thanks to all faculty members who've given their time and advice over the years. I want to thank the members of the Paschal lab for their dedication, advice, and support. Special thanks to former members Adam Spencer, Chelsi Snow, Leonard Shank, Natalia Dworak, Dawid Makosa and Kasey Jividen who helped form both foundations of some of the work done in this study and a great lab environment. Thanks to current members, Teddy Kamata, Bart Remlein, and Chun-Song Yang, for being great scientists. I have also had a second home in the McConnell lab during my time in Virginia, thank you to past and present lab members for all you've done.

Thank you to the members of my committee: Bettina Winckler, Mazhar Adli, and Mark Beenhakker. Your input has been much appreciated. Special thanks to Mike McConnell who has basically been a second mentor to me and given me his time, support, expertise, a place in his lab, and demonstrated his enthusiasm for the work we do.

Thanks to my advisor Bryce Paschal for all he has done for me. The relationship between mentee and mentor can make or break the graduate experience, and I feel that ours has been a great one which has helped me develop into a careful, independent scientist.

Finally, I would like to thank my family for all the encouragement they have given throughout my career. My Mom, Sheila, for her compassion and willingness to be where she is needed. My Dad, Tim, for introducing me to science and showing me the meaning of work ethic. My amazing siblings for always making life enjoyable. And finally, all my extended family for your love and support.

Table of Contents

Publication Information

The basis of Chapter IV of this study was the following publication:

Oostdyk, L.T., Shank, L., Jividen, K., Dworak, N., Sherman, N.E., and Paschal, B.M. (2019). Towards improving proximity labeling by the biotin ligase BirA. Methods *157*, 66–79.

List of Abbreviations

ADP	Adenosine diphosphate
AMP	Adenosine monophosphate
ANOVA	Analysis of Variance
APEX	Engineered ascorbate peroxidase
ARM	Armadillo repeat
ATP	Adenosine triphosphate
Autorad	Autoradiography
BAP	Biotin acceptor peptide
BCCP	Biotin carboxyl carrier protein
bioAMP	Biotinoyl-5′-AMP
BioID	Proximity dependent biotin identification
bNLS	Bipartite nuclear localization signal
BSA	Bovine serum albumin
CAS	Exportin-2
CB	Coomassie blue
cNLS	Classical nuclear localization signal
DDB2	DNA damage-binding protein 2
CTCF	CCCTC-binding protein
DNA	Deoxyribonucleic acid
dHSF	Drosophila heat shock factor
ESC	Embryonic stem cell
FG repeat	Phenylalanine-glycine repeat
fl-neutra	Fluorescent neutravidin
FP	Fluorescence polarization anisotropy
GDP	Guanosine-5’-diphosphate
GFP	Green fluorescent protein
GST	Glutathione S-transferase
GTP	Guanosine-5'-triphosphate
hnRNP	heterogeneous nuclear ribonucleoprotein
IBB	Importin β binding domain
IBB2	KPNA2 IBB
IBB7	KPNA7 IBB
IF	Immunofluorescence
Imp-α	Importin α
Imp-β	Importin β
IP	Immunoprecipitation
iPSC	human induced pluripotent stem cell
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IVBB	In vitro biotinylation buffer
K_D	Eqilibrium dissociation constant
kD	Kilodalton

KPNA	Karyopherin α
LC-MS	Liquid chromatography mass spectrometry
MBP	Maltose binding protein
mNLS	Monopartite nuclear localization signal
mRNA	Messenger RNA
NES	Nuclear export signal
NLS	Nuclear localization signal
NPC	Nuclear pore complex or neural progenitor cell
NTF2	Nuclear transport factor 2
NUP	Nucleoporin
POI	Protein of interest
PPI	Protein–protein interaction
Rb	Retinoblastoma protein
RCC1	Regulator of chromosome condensation 1
RGG	Arg-Gly-Gly binding motif
RNA	Ribonucleic acid
RRM	RNA recognition motif
SEC	Size exclusion chromatography
shRNA	Short hairpin RNA
SILAC	Stable isotope labeling by amino acids in cell culture
siRNA	Short interfering RNA
SV40	simian virus 40
TBS	tris buffered saline
TLC	thin layer chromatography

Chapter I

Introduction

I. The Essential Function of Nuclear Transport

In eukaryotic cells, genetic material is sequestered inside the nucleus, a unique, double membrane-bound organelle. Due to this fact, the control of gene expression and DNA replication is isolated from the rest of the cell, including the ribosome where protein translation occurs. This arrangement necessitates the export of certain materials generated inside the nucleus, such as mRNAs, to the cytoplasm, and the import of proteins with nuclear functions. To permit this bi-directional transport, the double-lipid bilayer of the nuclear envelope contains large macromolecular assemblies termed nuclear pore complexes (NPCs). The massive NPC structure is approximately 120 MDa in humans and integrates between 500 and 1000 individual copies of about 30 different proteins (nucleoporins; NUPs) into a pore structure with 8-fold rotational symmetry. At the center of this pore structure are NUPs which contain a meshwork of phenylalanine-glycine (FG) repeats that help control passage through the pore (Tran and Wente, 2006).

Bi-directional transport through the nuclear pore can occur in two distinct manners, passive diffusion and active transport. Molecules generally less than 40-60 kDa (~40 nm) in size, such as salts, nucleotides and small proteins, can translocate by simple passive diffusion (Panté and Kann, 2002). Molecules larger than 40-60 kDa, require active transport through the pore which is mediated by the recognition of a transport signal by a diverse family of soluble nuclear transport receptors (Mattaj and Englmeier, 1998; Weis, 2003). Furthermore, even molecules which can freely diffuse through the nuclear pore may require active transport to permit accumulation against a concentration gradient. Active transport, by nature then, requires energy to facilitate translocation. For the largest class of transport receptors, this energy comes in the form of the small GTPase, Ran, and an asymmetric distribution of its two nucleotide states between the cytoplasm and nucleus (Melchior et al., 1993; Moore and Blobel, 1993). As a GTPase, Ran can cycle between the GTP and GDP bound forms. By maintaining a high concentration of RanGTP in the nucleus and RanGDP in the cytoplasm, an energy gradient is created between the two cellular compartments (Kalab et al., 2002). The loading of Ran with GTP is catalyzed specifically in the nucleus by its guanine nucleotide exchange factor, RCC1, while GTP hydrolysis occurs specifically in the cytoplasm and is mediated by RanGAP and its cofactor RanBP1 (**Fig 1.1A**) (Bischoff et al., 1994, 1995; Izaurralde et

al., 1997). Furthermore, not only is there spatial separation between the two forms of Ran, but also an asymmetrical distribution of the total Ran protein in the cell. RanGDP is selectively shuttled into the nucleus by NTF2 at a higher rate than it is exported giving rise to what is called the Ran gradient (Paschal and Gerace, 1995; Ribbeck et al., 1998; Smith et al., 1998). Together, the gradient of total Ran protein and the separation of the GTP and GDP bound forms provides the energy necessary for classical nuclear transport to occur against cargo concentration gradients both into and out of the nucleus.

II. Mechanisms of Transport

A. *The Nuclear Localization Signal – a ticket to the nucleus*

The classical manner by which proteins with nuclear functions are marked for import is by the presence of a short nuclear localization signal (NLS). The classical NLS (cNLS) contains either one (monopartite) or two (bipartite) clusters of basic amino acid residues which are specifically recognized by the nuclear import machinery. The canonical monopartite cNLS comes from the SV40 large T-antigen (^{126}PKKKRKV^{132}) (Kalderon et al., 1984), while the canonical bipartite cNLS comes from Nucleoplasmin (^{155}KRPAATKKAGQAKKKKL^{170}) (Dingwall and Robbins, 1988; Robbins et al., 1991). Further studies have determined many key requirements for each of the two types of cNLS and identified loose canonical sequences. Monopartite cNLS loosely follow the pattern K(K/R)*X*(K/R) (Hodel et al., 2001), while bipartite cNLS fit KR$X_{10–12}$K(K/R)(K/R) (Robbins et al., 1991). The variability in both mono- and bipartite cNLS amino acid sequences modulates the affinity with which they bind to nuclear transport machinery as well as the rate and specificity of nuclear transport, which we will discuss in further detail below. Moreover, non-classical NLS have also been described which differ from the consensus sequences above (Kosugi et al., 2009a). In this thesis, the term NLS will refer to classical NLS sequences unless otherwise noted. The determination of a functional NLS is made by demonstrating both sufficiency and necessity of the sequence for directing nuclear translocation of a given protein (Damelin et al., 2002). This includes targeting an unrelated protein to the nucleus simply by the addition of the NLS and disrupting nuclear localization by mutation or removal of the NLS. The transport of NLS containing proteins is carried out by importins (also called karyopherins). This includes

importin α (Imp-α), which acts as an adapter protein to bind the NLS, and Importin β (Imp-β), which binds to Imp-α and facilitates transport through the NPC.

B. *Importin β – the master transporter*

Imp-β (also called karyopherin β1) is the founding member of the importin β superfamily of nuclear receptors which contains more than 20 members in humans (Harel and Forbes, 2004; Mosammaparast and Pemberton, 2004; Ström and Weis, 2001). Importin β family members vary in size between 90 and 130 kDa and share low sequence identity (15-20%) but high structural similarity (O'Reilly et al., 2011). All are made up of varying numbers of HEAT repeats, typically around 20. HEAT repeats are helix-loop-helix motifs which stack in Importin β family members to allow the overall protein to coil into a super-

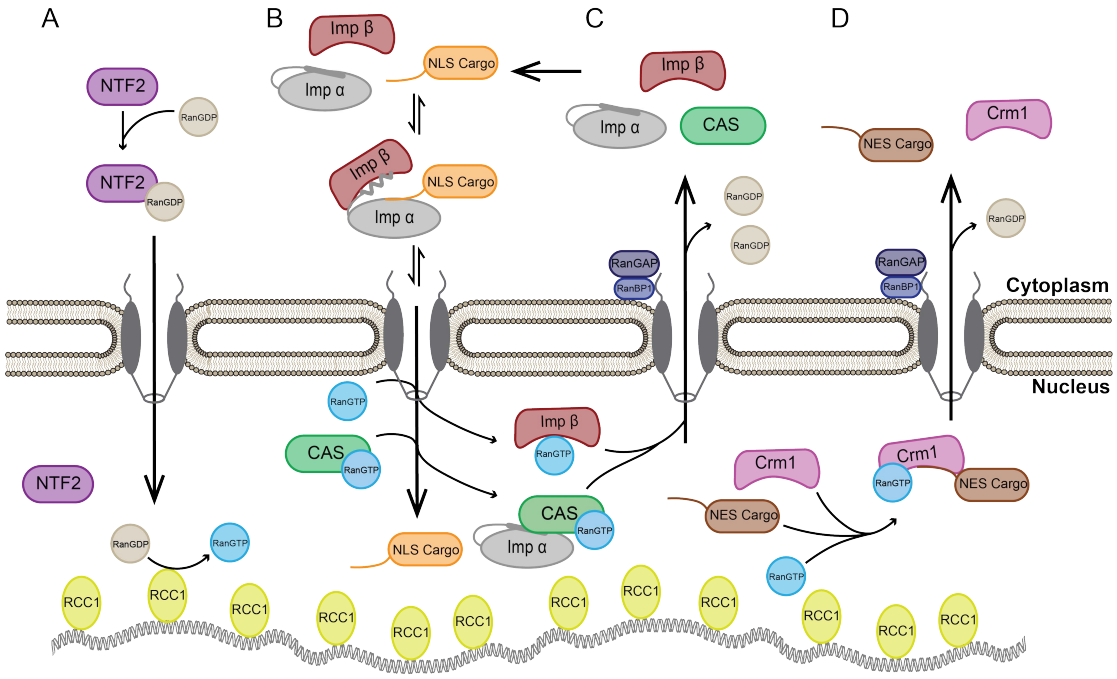


Figure 1.1. The nuclear transport cycle.
The nuclear transport cycle is driven by the Ran gradient. A. The nuclear import of RanGDP is facilitated by the transport factor NTF2. In the nucleus, the chromatin associated guanidine nucleotide exchange factor RCC1 converts RanGDP to RanGTP to maintain the gradient of the two states of Ran. B. Classical nuclear import is mediated by Imp-β and Imp-α which form a heterotrimeric complex with an NLS containing cargo protein. The complex dissociates in the presence of RanGTP in the nucleus. C. Imp-β and Imp-α are recycled back to the cytoplasm. Imp-α recycling is mediated by the export factor CAS. In the cytoplasm RanGAP facilitates hydrolysis of RanGTP to RanGDP to dissociate the complex. D. Nuclear export of NES containing proteins is facilitated by the export factor Crm1, which binds to the NES in the presence of RanGTP.

helical structure (Cingolani et al., 1999a; Vetter et al., 1999). Members of the Importin β

family function as either nuclear import or export factors, or sometimes both. Imp-β, which acts solely as an import factor, can directly bind NLS sequences or bind to an NLS-like region in the adaptor-receptor Imp-α, and facilitate nuclear translocation of the heterodimeric or heterotrimeric complex (**Fig 1.1B**). The process of transport utilizing both Imp-β and Imp-α is termed classical nuclear transport.

C. *Importin α – the NLS adaptor*

The importin α family (also called karyopherin α, or KPNA, family) of receptors consists of 7 members in humans (Kelley et al., 2010; Pumroy and Cingolani, 2015). These receptors share a common structure which consists of an N-terminal Importin β binding (IBB) domain and 10 ARM repeats, which form the body of the receptor. An ARM repeat is a three-helix motif made up of approximately 40 amino acids, named for the Drosophila protein they were first identified in, Armadillo (Peifer et al., 1994). The series of ARM repeats in Imp-α form a super-helical structure which contains a shallow, concave binding surface along its length that is separated into major and minor NLS binding grooves (Conti et al., 1998). The major NLS binding groove is toward the N-terminus of the receptor, between the second and fourth ARM repeats, and recognizes both the entire basic amino acid cluster of monopartite NLS and one of the two basic clusters of bipartite NLS. The minor NLS binding groove is between the seventh and eighth ARM repeats and is primarily used in the binding of one of the two elements of the bipartite NLS (Conti et al., 1998). Despite the small size of the NLS, the binding grooves of Imp-α form extensive contacts with NLS sequences on the basis of both charge and hydrophobicity (Conti and Kuriyan, 2000; Fontes et al., 2000, 2003). This multitude of interactions permits both high specificity and variability in NLS cargo binding.

On its own, Imp-α has low affinity for NLS cargo (~500-1000 nM) (Harreman et al., 2003a). This is due to the auto-inhibitory nature of the N-terminal IBB. This domain features two clusters of basic amino acids which mimic a bipartite NLS and fold back into the NLS binding groove in the absence of Imp-β (Kobe, 1999). We refer to this conformation of Imp-α as being in a ‘closed’ state. In the presence of both Imp-β and an NLS, however, the aptly named IBB (Imp-β binding) domain is displaced from the NLS

binding groove of Imp-α and engages with Imp-β, forming a stable heterotrimeric complex (Cingolani et al., 1999a). This is referred to as the 'open' state of Imp-α and results in a large increase in the affinity of Imp-α for NLS cargo (to ~10-20 nM) (Harreman et al., 2003a). Conversely, also due to auto-inhibitory interactions between the IBB and the NLS binding groove, the affinity of Imp-α for Imp-β is weak in the absence of an NLS cargo (Harreman et al., 2003a; Kelley et al., 2010). While an isolated IBB domain binds Imp-β with low nanomolar affinity (~17 nM), the affinity of full-length Imp-α for Imp-β in the absence of an NLS cargo is significantly lower (~600 nM) (Harreman et al., 2003a). Together, these characteristics suggest a cooperative mechanism of formation of the heterotrimeric transport complex which can only form stably when Imp-β, Imp-α and an NLS cargo protein are present.

D. *Nuclear translocation and recycling of the import complex*

After the formation of the heterotrimeric import complex, translocation through the nuclear pore is mediated by Imp-β (**Fig 1.1B**). This occurs through direct, primarily hydrophobic, interaction of two regions of Imp-β, one toward the N-terminus and one toward the C-terminus, with FG repeats in the NPC (Bayliss et al., 2000; Bednenko et al., 2003). Evidence suggests that the different NUPs and their FG repeats which make up the central channel of the NPC form an affinity gradient across the pore to help usher Imp-β into the nucleus (Ben-Efraim and Gerace, 2001; Pyhtila and Rexach, 2003). At the final FG NUP, Imp-β binds tightly unless released by RanGTP (Görlich et al., 1995) and transport can be aborted if RanGTP is not encountered (Lowe et al., 2010). Binding of RanGTP to Imp-β induces a conformational change and causes the release of the Imp-α IBB domain (Cook et al., 2007; Görlich et al., 1996). Release of the NLS cargo from Imp-α occurs via the concerted effort of the IBB domain (Harreman et al., 2003a), an autoinhibitory-like sequence in NUP50 (Gilchrist and Rexach, 2003), and the binding of the export factor CAS with RanGTP (Sun et al., 2008). In different studies, each of these steps has been shown to be necessary for efficient release of NLS cargo in the nucleus.

Once the NLS cargo is released, Imp-β and Imp-α must be recycled back to the cytoplasm in order to participate in further rounds of transport (**Fig 1.1C**). This process is also driven by the Ran gradient. In the RanGTP bound state, Imp-β moves back through

the NPC and associates with the cytoplasmic side of the NPC before being released by the hydrolysis of RanGTP to RanGDP by RanGAP and RanBP1. Export of Imp-α is mediated by CAS and RanGTP. CAS is a member of the Importin β family which acts as an export factor and only binds its cargo, Imp-α, in the presence of RanGTP. The crystal structure of the Imp-α:CAS:RanGTP complex has been solved and shows that CAS binds strongly (low nanomolar affinity) to the C-terminus of Imp-α and makes contacts with looping regions of the IBB domain (Köhler et al., 1999; Matsuura and Stewart, 2004). Critical to this interaction is the 'closed', auto-inhibited state of the IBB domain which ensures Imp-α cannot be recycled until after the NLS cargo has been released. Following the formation of the export complex, CAS interacts with the NPC to mediate cytoplasmic translocation before RanGTP hydrolysis in the cytoplasm induces dissociation and releases Imp-α, freeing it for another round of transport (Cook et al., 2005).

E. *Nuclear Export Signals*

The NLS is not the only signal sequence which directs proteins to pass through the nuclear pore complex. The nuclear export signal (NES) is a leucine-rich signal which directs proteins to be transported from inside the nucleus back to the cytoplasm (Fornerod et al., 1997; Fukuda et al., 1997; Ossareh-Nazari et al., 1997; Stade et al., 1997). Crm1, another member of the importin β superfamily, binds to NES containing proteins in the presence of RanGTP in the nucleus and facilitates export through the NPC to the cytoplasm (**Fig 1.1D**).

III. Diversity of Importin α

F. *Importin α isoforms*

Prior to this point in the text, the importin α family in humans has been referred to by the singular name, Imp-α. It was previously mentioned, however, the importin α family is made up of seven members in humans. This is compared to a single Imp-α in yeast, three isoforms in Drosophila and six isoforms in mice. Having multiple isoforms allows for some diversity in their functions, primarily as it relates to NLS cargo specificity and expression levels. Before further discussion, the naming convention of the Importin α family must be addressed. Over the years, differences in the naming of importin α isoforms by authors in the field have led to some confusion in the literature. Importin α

isoforms are named both as Imp-αX (typically for the protein) as well as KPNAX (typically for the gene), and the numbers for each isoform do not necessarily agree. For example, the prototypical importin α is named both Imp-α1 and KPNA2 which can easily be confused with another importin α isoform which is named both Imp-α5 and KPNA1. Furthermore, nearly all of the isoforms have additional names which were given before the unification of either the importin or karyopherin nomenclature. The naming conventions for each Importin α isoform is summarized in Table 1.1. In this text, the KPNA naming convention will be used when referring to both each individual protein and gene, while importin α (Imp-α) may be used to refer to the family as a whole.

Table 1.1 Naming conventions of the Karyopherin α family of receptors.

KPNA	Importin α	Gene ID	Accession	Aliases
KPNA1	Importin α5	3836	CAG33024.1	IPOA5, NPI-1, RCH2, SRP1
KPNA2	Importin α1	3838	NP_001307540.1	IPOA1, QIP2, RCH1, SRP1-alpha
KPNA3	Importin α4	3839	NP_002258.2	IPOA4, SRP1gamma, SRP4, hSRP1
KPNA4	Importin α3	3840	AAH16754.1	IPOA3, QIP1, SRP3
KPNA5	Importin α6	3841	NP_001353235.1	IPOA6, SRP6
KPNA6	Importin α7	23633	CCQ43410.1	IPOA7
KPNA7	Importin α8	402569	NP_001139187.1	IPOA8

Table 1.1 The human KPNA family naming conventions.

The seven human KPNA isoforms share between 41 and 86% identity in their amino acid sequences overall (Kelley et al., 2010; Pumroy and Cingolani, 2015). The isoforms are grouped into three subfamilies, α1, α2, and α3, based on their levels of similarity (**Fig 1.2A**). The α1 subfamily contains KPNA1, KPNA5, and KPNA6 which share a minimum of 81% identity. In the α3 family, KPNA3 and KPNA4 are 86% identical. The α2 subfamily, which contains the prototypical KPNA2, as well as the most recently discovered importin α, KPNA7, shares only 55% identity between the two members. Despite the variations in primary sequence, each Imp-α isoform shares the common structure of an N-terminal IBB domain and body made up of ARM repeats, with major and minor NLS binding grooves (Miyamoto et al., 2016). Additionally, a much higher level of amino acid conservation is observed along the entirety of the NLS binding groove across all seven isoforms (**Fig 1.2B**) (Kelley et al., 2010).

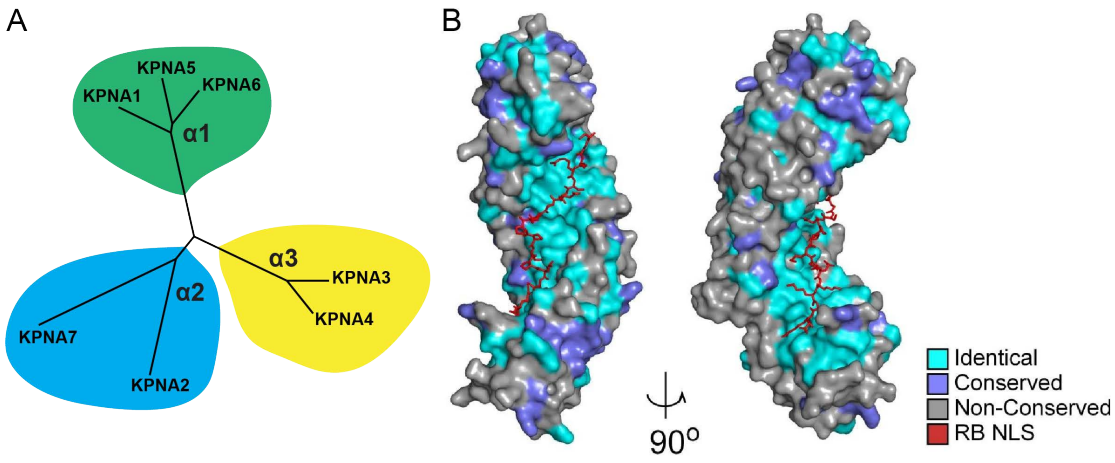


Figure 1.2: The human Importin α family.
A. Phylogenetic tree of the seven human Imp-α isoforms. The seven isoforms cluster into three subfamilies (α1, α2 and α3) which are labeled in green, blue, and yellow respectively. B. Structural model illustrating the conservation of amino acids between the seven Imp-α isoforms. Residues are highlighted in the indicated colors to show identical and conserved residues between all seven isoforms. The bipartite NLS from retinoblastoma protein (Rb) is shown in red. PDB: 1PJM. Rendered in Pymol. Figure adapted from (Kelley et al., 2010).

G. *Importin α cargo specificity*

Despite the high conservation of the NLS binding groove, there are differences in the affinity of each Imp-α isoform for certain NLS sequences. Accordingly, there are both shared and distinct NLS cargo proteins between the isoforms (Pumroy and Cingolani, 2015). For example, the α3 subfamily (KPNA3 and KPNA4) was shown to exclusively facilitate the nuclear import of the critical RanGEF, RCC1 (Köhler et al., 1999). NLS cargo specificity of Imp-α isoforms and the resulting differences in transport capacity is utilized by the cell to affect many cellular processes. Multiple examples will be discussed in the next section.

H. *Importin α isoform switching*

Expression of importin α isoforms varies significantly between different cell and tissue types and over different developmental programs (Köhler et al., 1997; Pumroy and Cingolani, 2015; Tsuji et al., 1997). This allows cells to alter their transport capacity, both generally and for specific cargoes, on the basis of which Imp-α isoforms are being expressed (Goldfarb et al., 2004; Pumroy and Cingolani, 2015; Yasuhara et al., 2009). Developmental programs such as oogenesis, spermatogenesis, and embryonic

development rely heavily on differential expression of Imp-α isoforms (Hogarth et al., 2006; Holt et al., 2007; Mason et al., 2002; Miyamoto et al., 2012).

i. *Drosophila melanogaster and Caenorhabditis elegans*

In *D. melanogaster*, which has three Imp-α isoforms, there are specific roles for the α2 isoform during oogenesis and spermatogenesis (Mason et al., 2002). Interestingly, while all Drosophila Imp-α2 mutant flies are sterile, male, but not female, fertility can be recovered by overexpression of one of the other two Imp-α isoforms. This indicates that Drosophila Imp-α2 has an essential function in oogenesis, but a non-essential function in spermatogenesis (Mason et al., 2002). During embryogenesis, Imp-α3 is not expressed early in development, but from cell cycle 13 onwards it is expressed and can facilitate nuclear translocation of the Drosophila heat shock factor (dHSF) in response to heat shock (Wang et al., 1998). In fact, dHSF is expressed earlier in development, but the embryo cannot respond to heat shock until later developmental timepoints when Imp-α3 is expressed.

The three Imp-α isoforms in *C. elegans* are regulated in a similar manner. While all three isoforms are highly expressed in the adult worm, Imp-α3 is the dominant isoform in the developing embryo. RNAi knockdown of Imp-α3 in the developing worm showed Imp-α3 is required for the progression of meiotic prophase I during oocyte development. Furthermore, Imp-α3 knockdown worms display a mislocalization of nucleoporins suggesting a role of Imp-α3 which is distinct from the other isoforms in all cells (Geles and Adam, 2001).

ii. *Mus musculus*

Mice have six Imp-α isoforms which have been studied extensively in a multitude of different developmental programs. Distinct regulation of Imp-α isoforms has been reported in mouse germ cell growth and development, including both oogenesis (Mihalas et al., 2015) and spermatogenesis (Hogarth et al., 2006; Holt et al., 2007). During oogenesis, there are significant changes in Imp-α RNA levels, as well as changes in protein levels and subcellular localization (Hogarth et al., 2007; Mihalas et al., 2015). All six Imp-α isoforms are expressed in the embryonic ovary, however, in postnatal folliculogenesis Kpna1, Kpna2, Kpna4, and Kpna6 are significantly upregulated and

Kpna7 is down-regulated during the primordial to primary follicle transition. Similar changes are observed in spermatogenesis, with each of the mouse Imp-α isoforms being detected in adult testis. A careful study of the localization and expression patterns of Imp-α isoforms in post-natal and adult mice revealed distinct patterns of expression throughout (Hosokawa et al., 2008). Kpna1, 4 and 6 were expressed throughout the brain and spinal cord, while Kpna2 was lowly expressed in these regions but was moderately expressed in the olfactory bulb and reticular system.

A specific example of Imp-α isoform switching in mice was demonstrated in embryonic stem cells (mESCs) and concerns the process of neuronal differentiation (Yasuhara et al., 2006, 2009). The Yoneda group observed that in mESCs expression of Kpna2 is high, while Kpna1 and Kpna4 are both expressed lowly, or not at all. After inducing neuronal differentiation with retinoic acid, expression of Kpna2 is lost but Kpna1 and Kpna4 are upregulated. These expression changes were found to not just be correlated with neuronal differentiation, but actually sufficient to drive differentiation toward a neuronal state in their own right. Specifically, both siRNA-mediated knockdown of Kpna2 and exogenous expression of Kpna1 induced expression of neuronal genes such as Nestin and Map2. These changes were found to be mediated, at least in part, by transport of a few key NLS containing cargo proteins: Oct3/4, Sox2, Oct6, and Brn2. While both receptors were able to facilitate import of the pro-pluripotency factors Oct3/4 and Sox2, only Kpna1 could facilitate import of the pro-neural factors Oct6 and Brn2. In fact, in a subsequent study, the authors determined that Kpna2 could promote cytoplasmic retention of Oct6 and Brn2 by binding the NLS of each protein at a non-classical location on the C-terminus of the transport factor (Yasuhara et al., 2013).

Interestingly, while Kpna1 expression is high in the adult mouse brain and Kpna2 expression is low, Kpna1 knockout mice did not display any abnormalities in brain development (Shmidt et al., 2007). In these mice, Kpna3 expression was upregulated specifically in the brain and, despite being from a different subfamily, appears to compensate for the necessary function of Kpna1. This result suggested that, *in vivo*, isoform switching of Imp-α isoforms can also occur to compensate for a lost isoform. In response to this study, the Yoneda group revisited the Kpna1 knockout mouse and

confirmed that the mice develop normally, however they identified abnormalities in the uterus and ovary of female mice and determined there was altered regulation of the progesterone receptor (Moriyama et al., 2011). These data suggest that while in some cases the Imp-α isoforms may be interchangeable (i.e. Kpna1 and Kpna3 in neuronal development), in other developmental programs this is not the case (such as in the female reproductive system).

In addition to Kpna1, knockout mouse models of other Imp-α isoforms have been used to study development in mice. Knockout of each of Kpna3 (Gabriel et al., 2011; Rother et al., 2011), Kpna6 (Gabriel et al., 2011; Rother et al., 2011), and Kpna7 (Hu et al., 2010) yields viable mice, with normal development of the brain and most other organs. For Kpna6 and Kpna7 however, there are clear effects on the reproductive capacity of female knockout mice. Both Kpna6 and Kpna7 are maternal effect genes, meaning they are expressed in the ovary and oocytes before zygotic genome activation (Thompson et al., 1998). Knockout of Kpna6 results in embryos that have a reduced ability to progress past the first cleavage after fertilization and completely arrest at the 2-cell stage (Rother et al., 2011). Kpna7 is interesting in that its expression is mainly limited to the ovary of adult mice, and, as a maternal effect gene, is highly expressed in the germinal vesicle oocyte (Hu et al., 2010). Expression is rapidly lost during embryogenesis and is nearly absent after the 8-cell stage of development. Knockout of Kpna7 does not result in a complete loss of viability as Kpna6 does, but instead reduces reproductivity and causes a sex imbalance by preferentially affecting female embryos (Hu et al., 2010).

Differential roles of Kpna2 and Kpna4 have also been identified in the differentiation of pluripotent mESCs to embryoid bodies (EBs), three-dimensional aggregates of stem cells that approximate early embryogenesis and have ectodermal, mesodermal and endodermal layers (Koike et al., 2007; Young et al., 2011). In agreement with the study from the Yoneda group, in mESCs Kpna2 facilitates the import of Oct3/4 to promote and maintain the pluripotent state of the cells. In contrast, Kpna4 adopts a primarily nuclear localization in mESCs but becomes more cytoplasmic upon induction of differentiation to allow the facilitation of import of NLS containing cargoes that promote differentiation and down-regulate Oct3/4 (Young et al., 2011).

Regulation of Imp-α isoform expression has been implicated in the oligodendrocyte differentiation in mice as well (Laitman et al., 2017). Oligodendrocyte differentiation induced differential expression of all 5 expressed isoforms (Kpna7 was not expressed in these cells). Interestingly, a differential requirement of Kpna4 was observed depending on the type of pro-myelinating factor used to induce differentiation. siRNA-mediated knockdown of Kpna4 prevented oligodendrocyte differentiation when using one type of pro-myelinating factor, ciliary neurotrophic factor, but not another, triiodothyronine (Laitman et al., 2017). This result highlights the potentially highly specific transport functions of Imp-α isoforms in different developmental processes.

iii. *Homo sapiens*

Early studies of importin α isoforms in humans identified distinct expression patterns in different tissues and cell types (Köhler et al., 1997, 1999, 2002). KPNA5 was found to be expressed only in the testis, while other imp-α isoforms were detected at some level in most tissue types (Köhler et al., 1997, 2002). While it is not always feasible to study human development in a similar manner as other species, RNA-seq datasets exist that allow for comparison of expression levels of the seven isoforms in different developmental contexts. Single-cell RNA-seq of human preimplantation embryos and embryonic stem cells demonstrated changes in expression of the seven isoforms as large as 40-fold from oocytes to blastocysts (Yan et al., 2013). Analysis of the CORTECON database from Fasano and colleagues, which features RNA-seq analysis of the differentiation of human embryonic stem cells (hESCs) to cortical neurons in culture, indicates isoform switching between KPNA2 and other KPNA proteins occurs in humans in a similar manner to mice (van de Leemput et al., 2014).

iv. *Other species*

Evidence of importin α subtype specificity and isoform switching have been identified in multiple other species as well. These include mammals, such as *Sus scrofa* (pig) (Wang et al., 2012) and *Bos taurus* (bovine) (Tejomurtula et al., 2009), and fish, such as *Oncorhynchus mykiss* (rainbow trout) (Wang et al., 2014). Each of these examples shows Imp-α isoform switching during early development similar to that which has been characterized in mice.

Together, the multitude of studies identifying spatiotemporal expression of Imp-α isoforms in species ranging from Drosophila and *C. elegans* all the way to humans strongly support the universality of Imp-α isoform switching. While there are multiple examples of cargo specificity between the Imp-α isoforms there is also redundancy and expression of one isoform can often compensate for the loss of another in vivo. In key developmental programs such as oogenesis, spermatogenesis, and early embryogenesis however, there are specific functions for multiple isoforms which are required.

I. *Specific Importin α functions in neurons*

Imps-α isoform switching is an important part of the progression of neuronal differentiation in mice. In addition to the role of Imp-α isoforms in selectively transporting specific pro-neural transcription factors, such as Brn2, other neuronal specific functions of Imp-α's have been identified. These functions are similar to those in many cell types, but aspects of neuronal morphology such as the presence of axons and dendrites result in additional roles for Imp-α isoforms in neurons. A recent review has proposed an Importin Code for neuronal transport which expands the canonical Imp-α functions to include those which transport signals from distal axons and dendrites and are not always reliant on a traditional NLS (Lever et al., 2015).

The nuclear transport machinery, and Imp-α isoforms specifically, functions in facilitating retrograde transport from both areas of axonal injury, and from synapses at the end of dendrites. Early work in the marine mollusk *Aplysia californica* discovered that some retrograde signals after axonal injury were proteins targeted by an NLS (Ambron and Walters, 1996; Schmied and Ambron, 1997). Subsequently, it was discovered that Imp-α/β are present in axons, far away from the cell body, and were responsible for this retrograde signaling processes in rat sciatic nerves (Hanz et al., 2003). This process was found to be controlled by upregulation of Imp-β protein in the axon in response to an injury. Perhaps surprisingly, *Imp-β* mRNA, but not protein, was constitutively present in axons and this regulation was at the level of translation. In contrast, multiple isoforms of Imp-α protein were found to be constitutively present in axons where they are complexed with the retrograde motor dynein. Western blot analysis of the rat sciatic nerve axon identified the presence of KPNA2, KPNA3, and KPNA6

(Hanz et al., 2003). While subsequent studies have confirmed the function of dynein and microtubules in enhancing nuclear import of certain NLS containing cargo proteins, the specific mechanism of interaction between Imp-α and dynein has not been elucidated (Moseley et al., 2007).

It was also determined that Imp-α/β help facilitate the nuclear translocation of the activated MAP kinase Erk1/2 in response to axonal injury (Perlson et al., 2005). This function is also dependent on an association of Imp-α with dynein and is believed to function through an interaction of phosphorylated Erk1/2 (pErk1/2) with Vimentin which binds directly to Imp-β. Together, this complex of dynein, Imp-α, Imp-β, and vimentin functions to facilitate the retrograde transport of pErk1/2 from the distal axon to the cell body and nucleus despite the fact that Erk1/2 do not contain an NLS.

Another specific example of retrograde transport signaling in neurons dependent on Imp-α is the nuclear import of phosphorylated STAT3 in response to axonal injury. STAT3 has been shown to interact directly with both KPNA1 and KPNA6 (Ma and Cao, 2006). KPNA1 knockdown in a model of axonal injury prevented the retrograde transport of STAT3 in response to injury. Injection of STAT3 into the cell body, however, recovered proper signaling (Ben-Yaakov et al., 2012). This suggests that KPNA1 is required for the retrograde transport function, via the interaction with dynein, but not for the nuclear import of STAT3, which can likely be mediated by KPNA6 in the cell body. Together, these bodies of work strongly support a direct interaction of Imp-α with dynein that facilitates translocation of cargo proteins to the cell body and nucleus. The actual mechanism of this interaction of Imp-α and dynein has, however, not been identified.

Retrograde signaling of synaptic communication from dendrites to the nucleus also requires Imp-α. Imp-α/β were found in distal neurites and synaptic compartments of *Aplasyia* and rodent neurons and accumulated in the nuclei of these neurons after synapse activity (Thompson et al., 2004). In *Aplysia*, stimuli which produce long, but not short-term facilitation trigger this nuclear accumulation. In rodent hippocampal neurons, activation of NMDA receptors, but not depolarization, triggered this translocation (Thompson et al., 2004). Martin and colleagues identified an anchoring of Imp-α to an NLS in the C-terminus of the NR1-1a subunit of the NMDA receptor. Upon activation of

the NMDA receptor, and specific phosphorylation events proximal to the bipartite NLS, this interaction is lost and allows Imp-α to facilitate nuclear import of NLS containing signaling proteins (Jeffrey et al., 2009).

J. *Importin α in disease*

As previously discussed, differential expression of Imp-α isoforms plays a large role in the regulation of many developmental programs. It stands to reason then that disruption of these tightly controlled expression profiles, or mutation in the receptors which alter their function, could impact cellular homeostasis and potentially result in disease. Indeed, aberrant expression and/or localization of Imp-α isoforms are associated with multiple disease pathologies including many cancers and neurological disorders.

i. *Importin α in cancer*

The expression patterns of Imp-α isoforms are altered in multiple types of cancers. The most common alteration, which is generally a marker of poor prognosis, is the upregulation of *KPNA2*. In fact, studies have used KPNA2 as a biomarker initially in melanoma (Winnepenninckx et al., 2006) and breast cancer (Dahl et al., 2006), and subsequently in various cancers including cervical (van der Watt et al., 2009), esophageal squamous cell carcinoma (Sakai et al., 2010), ovarian (Zheng et al., 2010), lung (Wang et al., 2011), hepatocellular (Yoshitake et al., 2011), bladder (Jensen et al., 2011), prostate (Mortezavi et al., 2011), astrocyte glioma (Gousias et al., 2012), anaplastic oligoastrocytoma (Gousias et al., 2014), gallbladder (Xiang et al., 2019) and colorectal (Wang et al., 2019). In these studies, patients with high *KPNA2* expression show significantly shorter overall and disease-free survival, and it is likely this trend can be observed in other cancer types as well. Whether *KPNA2* expression itself is a driver or result of cancer progression varies between studies. Reduction of *KPNA2* levels by siRNA in prostate, hepatocellular carcinoma, gallbladder, and lung cancer cell lines reduced cell growth and migration (Mortezavi et al., 2011; Wang et al., 2011; Xiang et al., 2019; Yoshitake et al., 2011). A xenograft study of ovarian cancer tumors transplanted into mice showed increased tumorigenicity of tumors with high *KPNA2* expression and reduced tumorigenicity with siRNA-mediated *KPNA2* knockdown (Huang et al., 2013). In cervical cancer however, only the knockdown of other

components of the nuclear transport machinery Imp-β and Crm1, but not KPNA2 reduced cell proliferation, indicating different components of the transport machinery can be growth limiting in different cancer types (van der Watt et al., 2009).

The exact mechanism of increased *KPNA2* expression is not known, although one potential mechanism involves deregulated activity of the transcription factor E2F (Xiang et al., 2019). Additionally, increased expression of the deubiquitinase USP1 has been shown to regulate KPNA2 protein levels in breast cancer at the post-translational modification level. Finally, multiple mechanisms have been proposed for the role increased KPNA2 plays in promoting cancer progression; likely a reflection of the complex nature of cancer and the fact that active nuclear transport plays a role in so many different cellular processes.

There are fewer data linking other Imp-α isoforms to cancer. In fact, the only other isoform directly linked to cancer progression is KPNA7 in pancreatic cancer (Laurila et al., 2009, 2014). These studies will be discussed further later in the text. In addition, there is extensive literature, which links the modulation of nuclear export by the export factor Crm1 to cancer progression. Overexpression of Crm1 in some cancers acts to export tumor suppressing transcription factors such as p53 and BRCA1 (Mahipal and Malafa, 2016). In fact, multiple clinical trials are underway for the selective inhibitor of nuclear export, selinexor, which reduces Crm1 dependent nuclear export (Abdul Razak et al., 2016; Shafique et al., 2019).

ii. *Importin α in neurological disorders*

The disruption of signaling from the synapse to the nucleus in neurons is a common cause of neurological disorders (reviewed in Benarroch, 2019 and Ferreira, 2019). As previously discussed, the Imp-α family members play important roles in the development of the nervous system and the brain and even function in neuron-specific roles of retrograde signaling. Studies have shown that in the absence of individual Imp-α isoforms, such as in Kpna1 knockout mice, other isoforms can compensate for the lost activity to permit proper neural development (Shmidt et al., 2007). Nonetheless, multiple examples exist where specific Imp-α isoform variants are associated with neurological disorder.

Specific variants of KPNA3 have been shown to be associated with schizophrenia (Morris et al., 2012; Wei and Hemmings, 2005; Zhang et al., 2006). These studies did not determine causality, but simply identified a correlation of SNPs in KPNA3 and other genes with the occurrence of schizophrenia in some populations. Reduced expression of KPNA4, as a result of a specific SNP in the *KPNA4* gene, is also observed in schizophrenia (Roussos et al., 2013). Both KPNA3 and KPNA4 have been shown to facilitate the import of NF-κB (Fagerlund et al., 2005). NF-κB is constitutively active in neurons where it promotes neurite outgrowth and plays a role in plasticity (Gutierrez and Davies, 2011). Decreased expression of KPNA4 in the superior temporal gyrus is proposed to result in decreased NF-κB signaling, which in turn promotes schizophrenia (Roussos et al., 2013). In Alzheimer's disease, aberrant localization of KPNA5 has been observed with sequestration to Hirano bodies of hippocampal neurons occurring (Lee et al., 2006). The specific cause of this localization change is unknown, but it may result in an increase in cytoplasmic p27 which is also observed in Alzheimer's (Ogawa et al., 2003).

The Imp-α isoform KPNA7 was the first to be described with specific mutations being apparently causal of a neurological disorder (Paciorkowski et al., 2014). Autosomal recessive, compound heterozygous mutations were identified in a sibling pair with severe neurodevelopmental defect and epilepsy. This study will be discussed further in the KPNA7 section of this introduction (Section IV).

iii. *Importin α in other diseases*

Additional disease states have been also been linked to aberrant regulation of expression of Imp-α isoforms. In inflammatory bowel disease and Crohn's disease overexpression of KPNA3 and KPNA4 is observed as a result of decreased expression of prohibin (Theiss et al., 2009). In a mechanism that is opposite to that observed in schizophrenia, increased KPNA3 and KPNA4 helps promote inflammation by increasing TNF-α mediated NF-κB signaling, which has been shown to be transported by the α3 subfamily (Fagerlund et al., 2005; Theiss et al., 2009). This contrast in the effects of Imp-α function between the digestive tract and the nervous system further highlights the importance of controlling the expression of Imp-α isoforms in different tissues. The loss of Imp-α expression in aged

myocardial microvascular epithelial cells (MMEC) has been implicated in aging-related decreases in angiogenesis which can impact cardiac function and other tissues (Ahluwalia et al., 2010). Decreased expression of KPNA2 and KPNA4 in aged MMECs is associated with the reduced nuclear import of the transcription factor hypoxia-inducible factor 1α (HIF1α) and decreased activation of the potent stimulator of angiogenesis vascular endothelial growth factor (VEGF) (Ahluwalia et al., 2010, 2014).

K. *Non-transport functions of Importin α*

In addition to the canonical transport functions of the Imp-α family, specific instances of Imp-α isoforms functioning outside of this process have is. In general, while each of these functions of Imp-α is outside nuclear transport they are primarily based on the binding of Imp-α to an NLS and therefore new adaptations of the normal protein function.

Nuclear transport factors play essential roles in mitosis, despite the fact that the nuclear membrane is disassembled and there is no need for nucleocytoplasmic transport (Forbes et al., 2015; Mosammaparast and Pemberton, 2004). The discovery that the transport machinery regulates mitotic spindle assembly was made with the observation that RanGTP addition promoted the assembly of microtubule asters in *Xenopus* mitotic egg extracts (Carazo-Salas et al., 1999). It was determined that RanGTP addition caused Imp-α/β to release spindle assembly factors (SAFs), including TPX2 (Target Protein for Xenopus Kinesin-like Protein 2), which were bound and sequestered in cytoplasmic areas away from chromatin (Gruss et al., 2001). The binding of Imp-α to TPX2 is via a short non-classical NLS which binds in the minor NLS binding groove (Giesecke and Stewart, 2010). During mitosis, the Ran-GEF RCC1 remains bound to chromatin and releases RanGTP, while RanGAP remains in the cytoplasmic areas to catalyze the conversion to RanGDP of any RanGTP which diffuses away from the chromatin (Kalab and Heald, 2008). The action of these two enzymes maintains the Ran gradient which is required for nucleocytoplasmic transport and ensures that SAFs, such as TPX2, are bound in the cytoplasmic areas and efficiently released from Imp-α/β in close proximity to chromatin to allow spindle formation and chromosome segregation.

After chromosome segregation occurs, the nuclear envelope must reassemble. Early studies determined Ran was required for nuclear assembly in *Xenopus* extract (Hetzer et al., 2000). In a manner similar to the regulation of spindle assembly, it was found that both Imp-α and Imp-β played a role by binding to so-called membrane assembly factors such as lamins and nuclear pore proteins (Forbes et al., 2015). Work from Adam and colleagues determined that Imp-α binds to an NLS in LaminB and that exogenous Imp-α can inhibit the formation of the nuclear envelope (Adam et al., 2008). The RanGTP cloud produced by chromatin-bound RCC1 again regulates the localization of the cargo release (Forbes et al., 2015).

Finally, Imp-α can also play direct roles in regulating DNA synthesis and gene expression. During apoptosis, the IBB domain of multiple Imp-α isoforms can be cleaved by caspases which generate ΔIBB Imp-α proteins. These truncated isoforms then bind to the MCM replication licensing factor and prevent new DNA synthesis (Kim and Lee, 2008). There is evidence that Imp-α can regulate gene expression as well. KPNA2 can localize to the promoter of the serine/threonine kinase *STK35* and cause upregulation in the expression of the gene (Yasuda et al., 2012). The exact mechanism of this function is not known but could potentially be mediated by KPNA2 binding to transcription factors.

IV. Importin α family member KPNA7

KPNA7 is the most recently identified importin α family member in humans, as well as multiple other species (Hu et al., 2010; Kelley et al., 2010; Tejomurtula et al., 2009). In each case, the coding sequence for KPNA7 was identified based on similarity to other Imp-α isoforms, specifically the common α2 subfamily member KPNA2. The human KPNA7 gene prediction was based on a protein BLAST of the human KPNA2 protein sequence (Kelley et al., 2010). Validation of the predicted sequence was performed in the LNCaP prostate cancer cell line via amplification of the cDNA of the predicted gene. As described in the previous section KPNA2 and KPNA7 share 54.7 percent sequence identity, which is less than the identity between the α1 and α3 subfamilies. Sequence analysis indicates that KPNA7 has homologs in all vertebrates. Interestingly, the IBB domain of KPNA7 was found to be more divergent from the ancestral yeast KPNA than the ARM core and while KPNA7 bound weakly to the prototypical NLS sequences from

SV40 and RB, it bound strongly to Imp-β (Kelley et al., 2010). The biochemical characteristics of KPNA7 protein, including the regulation of NLS cargo binding and auto-inhibition via the IBB domain are the basis of Chapter II of this thesis.

L. *A maternal-effect factor – KPNA7 expression and function*

KPNA7 expression is low, or even undetectable in most adult cell types, but has been shown to have essential roles in embryo development in multiple species including mice, pigs, cattle and trout (Hu et al., 2010; Tejomurtula et al., 2009; Wang et al., 2014, 2012). In mice, Kpna7 RNA expression is observed in the ovaries of adult mice, but not in the brain, heart, kidney, liver, pancreas, skeleton or testes (Hu et al., 2010). In addition, Kpna7 is highly expressed in the germinal vesicle and meiosis II oocyte, as well as the zygote but is virtually absent at the 8-cell stage of embryogenesis (Hu et al., 2010). This expression is indicative of a maternal-effect factor, a protein which is maternally stocked in the oocyte and required for activation of the zygotic genome after fertilization (Thompson et al., 1998). A Kpna7 knockout mouse displayed reduced reproductive function and a sex imbalance with preferential lethality in females in both heterozygous and homozygous Kpna7-mutant embryos. Parthenogenic studies of pre-implantation embryos also indicated a role for Kpna7 in reproduction. Homozygous Kpna7 mutant embryos lose control of cell-cycle progression with improper (rapid) progression through the first and second cell divisions and failure to reach the blastocyst stage. Finally, mutation of Kpna7 induced abnormal expression of chromatin-modifying enzymes in late two-cell embryos, including down regulation of Dppa2, Dppa4 and Piwil2, and upregulation of Hdac3. These changes resulted in a global down-regulation of the H3K27me3 mark which indicates a potential loss of heterochromatin and aberrant gene expression (Hu et al., 2010).

In cattle, similar regulation of KPNA7 expression and function was observed. *KPNA7* mRNA is detected at high levels in the adult ovary, but nearly absent in other adult tissues (Tejomurtula et al., 2009). In the germinal vesicle oocyte, *KPNA7* mRNA levels are ≥30-fold that of other Imp-α isoforms. Furthermore, expression of KPNA7 protein was found to be high in the germinal vesicle oocyte and persist to the 16-cell embryo but drop sharply in the blastocyst. The function of KPNA7 in the early embryo development

of bovine embryos was investigated by siRNA mediated knockdown. Embryos injected with a specific siRNA targeting KPNA7 showed a significant reduction in progression to the blastocyst stage (Tejomurtula et al., 2009). A similar study in porcine embryos identified high *KPNA7* mRNA expression in the germinal vesicle oocyte which decreases by greater than 1000-fold by the blastocyst stage (Wang et al., 2012). Additionally, siRNA mediated knockdown of KPNA7 in porcine embryos leads to a developmental arrest and failure to progress to the blastocyst stage (Wang et al., 2012). Porcine KPNA7 was also shown to interact strongly with the pro-neural transcription factor Brn2, although KPNA7 knockdown in 4-cell stage embryos did not effect Brn2 localization (Li et al., 2015).

The unique expression pattern of Kpna7 is even conserved in *Oncorhynchus mykiss* (rainbow trout) (Wang et al., 2014). *Kpna7* mRNA is highly expressed in the egg and lowly expressed in the testes, but undetectable in most other tissues. In the ovaries, *Kpna7* is highly expressed and Kpna7 protein is easily detectable. During embryonic development, the highest levels of *Kpna7* mRNA are observed in the unfertilized egg with a gradual decrease occurring in early-stage embryos until 3 days post-fertilization and a sharp decline to a near undetectable level at 4 days post-fertilization (Wang et al., 2014). This is similar to the expression of mammalian KPNA7 and suggests a role in embryonic pathways. In addition, via a yeast two-hybrid screen of trout cDNA clones, Kpna7 was found to interact with the egg-specific factor Stl3 and another unknown gonad-specific protein (Wang et al., 2014).

While no studies of the function of KPNA7 in early embryonic development in humans have been performed, one can examine RNA-Seq data sets (previously discussed) to evaluate KPNA7 expression at these stages. In the RNA-Seq from Yan and colleagues, KPNA7 is the most highly expressed Imp-α isoform in the oocyte, expressed at nearly 50-fold higher levels than any other isoform (Yan et al., 2013). KPNA7 remains the most highly expressed isoform until the 8-cell embryo and is lost completely by the late blastocyst stage. Additionally, in a single-cell proteomics study of human oocytes, more KPNA7 peptides were identified than any other Imp-α isoform (Virant-Klun et al., 2016). Together these studies suggest, in multiple species from fish to mammals, KPNA7 plays

a critical role in early embryogenesis, is a maternal-effect factor, and participates in Imp-α isoform switching during this period. The exact biological contributions of KPNA7 protein during this period is still being investigated but it seems likely to reflect the nuclear import of factors with critical embryonic functions.

Studies have been performed to attempt to identify what specific NLS-containing cargo proteins KPNA7 transports. These studies have utilized either recombinant or exogenously expressed KPNA7, since low protein levels and poor antibody quality have prevented using this strategy with the immunoprecipitation of endogenous KPNA7. Using a recombinant GST-KPNA7 fusion the Yoneda group identified KPNA7 interacting proteins in HEK293T cells (Kimoto et al., 2015). Among the proteins identified there was an enrichment for those that fall into the GO biological process categories of RNA processing and metabolism. Specifically, the authors identified the DNA-damage response protein DDB2 and demonstrated binding of KPNA7 to an NLS near the N-terminus of the protein (Kimoto et al., 2015). Recombinant KPNA7 also facilitated nuclear translocation of GFP-DDB2 in permeabilized cell transport assays. The functional significance of the KPNA7-DDB2 interaction in cells is not clear since other Imp-α isoforms can also facilitate DDB2 nuclear translocation.

M. *KPNA7 in disease*

i. *Cancer*

While KPNA7 shows low expression in the majority of adult tissues, including the pancreas, expression has been shown to be increased in some pancreatic cancers as a result of gene duplication events (Laurila et al., 2009, 2014). The 7q22 amplicon, which includes KPNA7 is amplified in about 25% of pancreatic cancers and cell lines. Expression of KPNA7 is particularly high in the AsPC-1 and Hs700T pancreatic cancer cell lines. siRNA mediated knockdown of *KPNA7* in these cell lines results in inhibition of cell growth and colony formation, leads to a G1 cell-cycle arrest by induction of p21, and also induces autophagy (Laurila et al., 2014). Another study extended the effects of *KPNA7* knockdown to breast cancer lines with *KPNA7* gene duplication events (Vuorinen et al., 2018). In addition to the reduction in cell growth, defects in mitotic spindle formation and induction of abnormal nuclear morphology were observed

(Vuorinen et al., 2018). To attempt to identify the cargo proteins which mediate these effects, the Kallioniemi group used exogenously expressed, tagged KPNA7 to purify associated proteins from the pancreatic cancer cell lines AsPC-1 and Hs700T (Vuorinen et al., 2017). In this dataset, there was an enrichment in proteins involved in mRNA metabolism, as well as those associated with ribonucleoprotein complexes, and greater than 50% of the identified binding partners contain a predicted NLS. The nuclear localization of two binding partners, ZNF414 and MVP, was found to be dependent on KPNA7 for import, and siRNA-mediated knockdown of these proteins also resulted in cell cycle arrest (Vuorinen et al., 2017). Whether the functions of KPNA7 observed in these studies are universal, or functions gained as a result of the gene duplication and re-expression are not known, but they give valuable insight into potential protein function in different cellular contexts.

ii. *Epilepsy-Associated Mutations in KPNA7*

Autosomal recessive compound heterozygous mutations in KPNA7 have been found to be associated with severe neurodevelopmental defect in two siblings with infantile spasms and subsequent intractable epilepsy consistent with Lennox-Gastaut syndrome (Paciorkowski et al., 2014). Structurally, each patient had severe malformations of the cerebellum and the corpus callosum including cerebellar vermis hypoplasia and agenesis of the corpus callosum. Finally, each patient was non-verbal and non-ambulatory. Through whole exome sequencing of the affected patients, their parents, and unaffected siblings, compound heterozygous mutations in *KPNA7* were identified (**Fig 1.3A**). As can be seen in the pedigree each parent, with one of the two *KPNA7* mutations, was developmentally normal and able to live a normal life. Additionally, two male siblings did not inherit either mutated copy and were developmentally normal.

Both point mutations (c.1015C>G and c.1030G>C) occur in exon 7 of *KPNA7* and result in the amino acid substitutions Pro339Ala and Glu344Gln (**Fig 1.3B**). Each of these substitutions is in the 7th ARM repeat of KPNA7 which is proximal to the NLS binding groove of the receptor (**Fig 1.3C**). Glu344 is conserved across all seven human Imp-α isoforms and the equivalent mutation in KPNA2 reduces binding to the bipartite NLS from retinoblastoma protein (Rb) (Paciorkowski et al., 2014). These results led to the

hypothesis, and the basis of Chapter III of this current study, that KPNA7 has functions in neuronal development and a reduction in KPNA7 dependent nuclear import of key NLS containing cargo proteins is the basis for the severe neurodevelopmental defects observed in the affected patients.

While the two mutations described above were not identified in other databases of single nucleotide polymorphisms (SNPs) or clinical variants, they are not the only associations of KPNA7 with epilepsy and neurological defects. While no other studies have identified

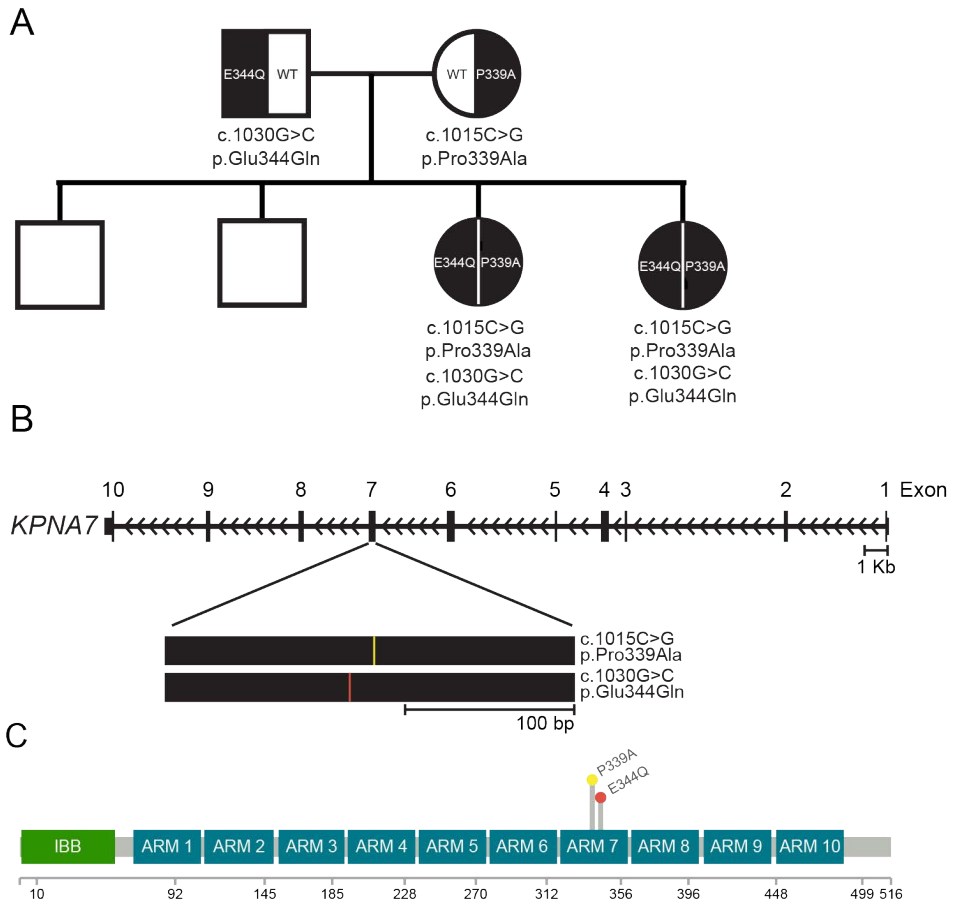


Figure 1.3 Epilepsy-associated mutation in KPNA7.
A. Pedigree of family with mutations in nuclear transport factor KPNA7. B. Diagram of the exon structure of the *KPNA7* gene. Epilepsy-associated mutations occur in exon 7. C. Protein diagram of KPNA7. Domains labeled include the Imp-β binding domain (IBB) and 10 Armadillo (ARM) repeats. Epilepsy-mutations in ARM 7 are indicated.

KPNA7 mutations as causal, multiple others have found mutations in the receptor in affected patients. A Leu203Phe substitution was identified in two Dravet syndrome patients (Kim et al., 2018b). A nonsense mutation, Arg36Ter, was identified by whole exome sequencing in a patient with mild intellectual disorder (Yang et al., 2013).

Finally, a Glu154Val substitution was identified in another whole genome sequencing study of a family with essential tremors (Odgerel et al., 2018). These associations further suggest a potential role for KPNA7 in neuronal development.

V. Concluding Remarks

The nuclear transport receptor is the least studied Imp-α isoform in humans. In the studies described in this thesis, we have investigated the biochemical characteristics of KPNA7 which make it distinct from other members of the Imp-α family of import receptors (Chapter II). We have then described functional changes imparted on KPNA7 protein and its expression by point mutations which are associated with neurodevelopmental disorder, discovered KPNA7 interacting proteins in neurons, and finally identified a regulatory mechanism of *KPNA7* gene expression in neurons (Chapter III). In a separate line of investigation, we have characterized how changes in a specific amino acid residue in the active site of the biotin protein ligase, BirA, alter its activity and use a new BirA mutant to identify protein-protein interactions in cells (Chapter IV).

Chapter II

Characterization of the Importin-β binding domain in nuclear import receptor KPNA7

I. Abstract

The Importin-α family of mammalian nuclear import receptors are encoded by seven genes that generate isoforms with 42-86% identity. Importin-α isoforms have the same protein architecture and share the basic property of NLS recognition, however, the tissue and developmental expression patterns of these receptors raise the question of what differences in Importin-α isoforms contribute to biological function. Here, we provide evidence that KPNA7, an Importin-α isoform with expression mostly limited to early development, adopts a conformation that is distinguishable from other family members. A highly conserved structural feature of nuclear import receptors (yeast to human) is the amino-terminal Importin-β binding (IBB) domain. The IBB domain occludes the NLS binding groove, which maintains the receptor in an auto-inhibited state prior to NLS recognition. In the presence of an NLS and Importin-β, which binds the IBB domain, auto-inhibition is relieved and Importin-α binds the NLS. Importin-β then targets the heterotrimer to the nuclear pore complex and mediates nuclear translocation. We demonstrate that KPNA7, unlike other Importin-α isoforms, shows evidence of an open state that permits constitutive binding to Importin-β in the absence of NLS. Importin-β contributes to KPNA7 activity, but it does so by enhancing NLS binding. Our data suggest the auto-inhibited state reflects features of both the IBB domain and the core structure of the receptor. Additionally, KPNA7 can maintain an open-state in the nucleus, suggesting it could also modulate the nuclear function of NLS-containing proteins.

II. Introduction

Nuclear import is responsible for facilitating the translocation of nuclear localization signal (NLS) containing proteins from the cytoplasm to the nucleus (Görlich and Kutay, 1999; Pemberton and Paschal, 2005). The protein sequences within NLSs are diverse but usually consist of either one (monopartite) or two (bipartite) clusters of positively charged, basic amino acid residues (Dingwall and Robbins, 1988; Kalderon et al., 1984). Classical nuclear import begins with the cytoplasmic assembly of a trimeric import complex, where Importin-α (Imp-α) binds to an NLS-containing protein cargo, and the NLS cargo-Imp-α binds to importin-β (Imp-β). The trimeric complex - composed of NLS cargo, Imp-α, and Imp-β - binds and translocates through the nuclear pore complex by virtue of Imp-β interactions with nucleoporins that line the central channel (Ben-Efraim and Gerace, 2001; Pyhtila and Rexach, 2003; Weis, 2003). Disassembly of the trimeric import complex occurs when it encounters nuclear Ran-GTP, which binds to Imp-β (Görlich et al., 1995, 1996).

S. cerevisiae contains a single Imp-α gene (Goldfarb et al., 2004), but in higher eukaryotes, Imp-α proteins are encoded by a multi-gene family, the number of which depends on the species. Imp-α proteins in higher eukaryotes can be divided into three distinct subfamilies, α1, α2, and α3 (Goldfarb et al., 2004; Mason et al., 2009; Miyamoto et al., 2012; Pumroy and Cingolani, 2015). There are three Imp-α isoforms in *C. elegans* and four isoforms in *D. melanogaster* (Mason et al., 2009). Mice have six isoforms and humans have seven isoforms, all encoded by separate genes (Pumroy and Cingolani, 2015). The founding member of the α1 subfamily, KPNA1 in humans, is most similar with yeast Imp-α, Srp1. The α2 and α3 subfamilies are thought to have evolved through duplication of the founding Imp-α and developed both redundant and cell and tissue specific roles that contribute to development and differentiation in higher eukaryotes (Goldfarb et al., 2004; Mason et al., 2009). Differential expression of Imp-α isoforms is a characteristic of all species with multiple isoforms, and temporal or tissue-specific expression often drives developmental programs. In *D. melanogaster* Imp-α2 is highly expressed in the early embryo, ovaries, and testes, and plays an essential role in oogenesis (Goldfarb et al., 2004; Mason et al., 2002). In mice, differential expression of

Imp-α5 (Kpna1) and Imp-α1 (Kpna2) helps drive the progression of neuronal differentiation (Yasuhara et al., 2006). The Imp-α isoforms are expressed in all adult cell types in humans with the exception of KPNA5, which is restricted to the testis (Köhler et al., 1997). While there is redundancy between Imp-α isoforms in terms of NLS selectivity, differences in Imp-α affinities for individual NLS cargoes could explain how differential Imp-α expression helps drive developmental programs (Pumroy and Cingolani, 2015). Thus, although nuclear import might be considered a housekeeping process, the aforementioned examples illustrate the important role of transport in nuclear regulation.

KPNA7 is the most divergent member of the Imp-α family in humans (Kelley et al., 2010). KPNA7 was assigned to the α2 subfamily based on 54.7% amino acid identity with KPNA2 and has homologs in all vertebrates. Protein expression of KPNA7 and its orthologs seems to be mostly limited to the oocyte and early embryogenesis. KPNA7 is the most highly expressed Imp-α isoform in germinal vesicle oocytes and the fertilized zygote, but KPNA7 expression plummets and is nearly absent in 8-cell embryos suggesting it could be a maternal effect gene (Thompson et al., 1998; Yan et al., 2013). Early developmental expression of KPNA7 has shown in multiple species including mice, pig, cows, and trout (Hu et al., 2010; Li et al., 2015; Tejomurtula et al., 2009; Wang et al., 2014, 2012). In mice, KPNA7 knockout compromises reproduction and causes sex imbalance, as most female embryos lacking KPNA7 are inviable (Hu et al., 2010). Analysis of pre-implantation embryos from these mice revealed an increase in cell cycle progression, aberrant control of cell division, abnormal cell morphology and down-regulation of H3K27me3 levels (Hu et al., 2010). In cows, siRNA mediated KPNA7 knock-down in early embryos results in a decreased proportion of embryos developing to the blastocyst stage (Tejomurtula et al., 2009).

With the exception of the ovary, KPNA7 expression in most adult cell types is extremely low. In certain pancreatic cancers, however, re-expression of KPNA7 can occur via gene duplication events (Laurila et al., 2009, 2014). Interestingly, in these cancers, KPNA7 promotes malignancy and was shown to be involved in the nuclear import of the nuclear proteins MVP and ZNF414 (Laurila et al., 2014; Vuorinen et al., 2017). Furthermore,

siRNA-mediated depletion of KPNA7 in pancreatic and breast cancer cell lines reduces cell proliferation and alters mitotic spindle formation, nuclear morphology, and the nuclear lamina (Vuorinen et al., 2018). In humans, KPNA7 function has been linked to brain function in a setting where compound-heterozygous mutations in KPNA7 are associated with severe neurodevelopmental defects (Paciorkowski et al., 2014).

An important regulatory feature of classical nuclear transport is the auto-inhibitory activity of the N-terminal ~60 amino acids of Imp-α, termed the importin-β binding (IBB) domain (Harreman et al., 2003a; Kobe, 1999). The IBB domain contains two clusters of positively charged, basic amino acid residues that mimic a bipartite NLS sequence. In the absence of an NLS cargo, the highly flexible IBB domain folds back onto the body of the receptor and occupies the NLS binding groove; this blocks NLS binding to Imp-α and prevents the IBB domain from binding to Imp-β in the absence of NLS cargo (Catimel et al., 2001; Conti and Kuriyan, 2000; Fanara et al., 2000; Harreman et al., 2003b; Kobe, 1999; Lott and Cingolani, 2011; Matsuura and Stewart, 2004; Rexach and Blobel, 1995). This represents the auto-inhibited, or ‘closed’ state, of Imp-α (Kobe, 1999). The IBB domain is conserved in S. cerevisiae and the auto-inhibition mechanism for Imp-α is essential (Harreman et al., 2003b, 2003a). When an NLS cargo and Imp-β are both present, the IBB domain is displaced from binding the NLS binding groove of Imp-α and binds to Imp-β with an α-helical structure (Cingolani et al., 1999b). This represents the uninhibited, or ‘open-state’, of Imp-α which is necessary for its transport function. After translocation through the nuclear pore, dissociation of Imp-β from the complex is induced by RanGTP, which binds Imp-β with sub-nanomolar affinity (Görlich et al., 1996). The NLS is displaced from Imp-α by the IBB domain in conjunction with NUP50, CAS, and RanGTP (Gilchrist et al., 2002; Sun et al., 2008). Recycling of Imp-α to the cytoplasm is mediated by binding to CAS and RanGTP, which involves CAS contact with a region of the IBB domain that is positioned for binding when Imp-α is in the auto-inhibited state (Kutay et al., 1997). Here, we present biochemical and cell biological evidence showing that KPNA7 adopts a predominantly open state independent of NLS binding. Removal of the IBB domain from KPNA7, which in other Imp-α proteins relieves auto-inhibition, results in a strong defect in NLS binding. Moreover, Imp-β binding to the IBB domain was found to enhance NLS binding

to KPNA7. Our data suggest the IBB domain can be used to modulate NLS binding through a mechanism that is separable from auto-inhibition. IBB domain swaps between KPNA2 and KPNA7 indicate the structural information that helps specify the open and closed states is contained within both the IBB domain and the body of the receptor.

III. Results

A. *Full-length KPNA7 binds Imp-β in the absence of NLS cargo*

The IBBs of Imp-α proteins expressed in humans (and other species, including *S. cerevisiae*) show a significant degree of sequence relatedness (Kelley et al., 2010), but the most conserved feature is the presence of two patches of amino acids, the KRR and RxxR motifs, that contact the major and minor NLS binding sites respectively (**Fig. 2.1A**). These arginine- and lysine-containing patches (**Fig. 2.1A**) contribute to the auto-inhibitory function mediated by the IBB because they occupy the same sites used by NLS cargo. This is illustrated in the crystal structures of full-length mouse Imp-α, yeast export complex-bound Imp-α, and yeast Imp-α with NLS, solved by the Kobe, Stewart and Kuriyan groups, respectively (Conti and Kuriyan, 2000; Kobe, 1999; Matsuura and Stewart, 2004) (**Fig. 2.1B**). Because Imp-β binding to the IBB of Imp-α proteins neutralizes the auto-inhibition mechanism, it indirectly stabilizes NLS binding (Cingolani et al., 1999b) (**Fig. 2.1B**). The major NLS binding sequence "KRR" is identical in all human Imp-α proteins, but there is some variation in the minor NLS binding sequence. The minor NLS binding sequence is "RQQR" in all eight KPNA7 proteins from mammalian species shown (**Fig. 2.1A**). This led us to consider whether variation in the minor NLS sequence or other features of the IBB in KPNA7 might be important for activity, possibly by affecting auto-inhibition.

The auto-inhibited state of Imp-α proteins can be assayed by Imp-β binding in the absence of NLS cargo. To this end, we expressed the seven human Imp-α paralogs as ^{35}S-methionine-labeled proteins by in vitro transcription and translation and performed binding assays with GST-Imp-β immobilized on glutathione beads. We observed robust binding of KPNA7 to Imp-β that was $\geq$4.5-fold higher than the level of binding of other Imp-α proteins (**Fig. 2.2A, B**). To determine if this reflects a direct interaction with Imp-

β, we combined purified, recombinant KPNA7 and Imp-β and analyzed complex formation by gel filtration chromatography. A standard curve was generated to determine apparent molecular weights of eluted species (**Fig. S2.1A, B).** MBP-KPNA7, which alone chromatographed as a monomeric species, eluted as a heterodimer when combined with Imp-β at a 1:1 molar ratio (**Fig. 2.2C, E)**. This indicates that KPNA7 can bind directly to Imp-β in the absence of other proteins.

B. *The IBB domain from KPNA2 can force KPNA7 to adopt a closed state*

A simple model to explain the differential binding of KPNA7 and KPNA2 to Imp-α is that the IBBs in these proteins have different affinities for Imp-β. Interrogation of the structure of the Imp-α IBB domain bound to Imp-β (shown in **Fig 2.1B**), and the residues in the IBB domain which mediate the interaction the crystal structure (highlighted in grey in **Fig 2.1A**) shows key residues are largely conserved, and any changes are within the same amino acid group. We compared the binding of the IBB domains from KPNA2 and KPNA7 to Imp-β by immobilizing each IBB as a GST protein on glutathione beads and performing binding assays using recombinant Imp-β. Binding profiles of Imp-β to the KPNA7 IBB and KPNA2 IBB were indistinguishable (**Fig. 2.2F, G**), suggesting the differences in Imp-α binding to Imp-β (**Fig. 2.2A, B**) is not explained by differences in affinity for the IBB. Another model to explain the differences in Imp-α binding to Imp-β is that most Imp-α isoforms adopt an auto-inhibited state wherein the IBB is inaccessible. By contrast, KPNA7 could adopt an open-state that is permissive for binding Imp-β. The extent to which an Imp-α protein adopts the auto-inhibited versus open-state in the absence of NLS cargo could involve features of the IBB, the NLS binding sites, and the positioning of Armadillo (ARM) repeats within the body of the receptor. To address the basis of auto-inhibition of KPNA2 and the open-state of KPNA7, we performed "IBB swaps" wherein the IBBs from KPNA2 and KPNA7 were deleted, the IBB from KPNA2 was fused to KPNA7, and the IBB from KPNA7 was fused to KPNA2. We expressed the full-length, deletion, and chimeric proteins (**Fig. 2.3A**) as ^{35}S-methionine labeled polypeptides and performed binding to GST-Imp-β. As expected, KPNA2 and KPNA7 lacking an IBB failed to bind Imp-β (**Fig. 2.3B**). Remarkably, transplanting the KPNA2 IBB onto KPNA7 was sufficient to reduce Imp-β binding to a level comparable to native

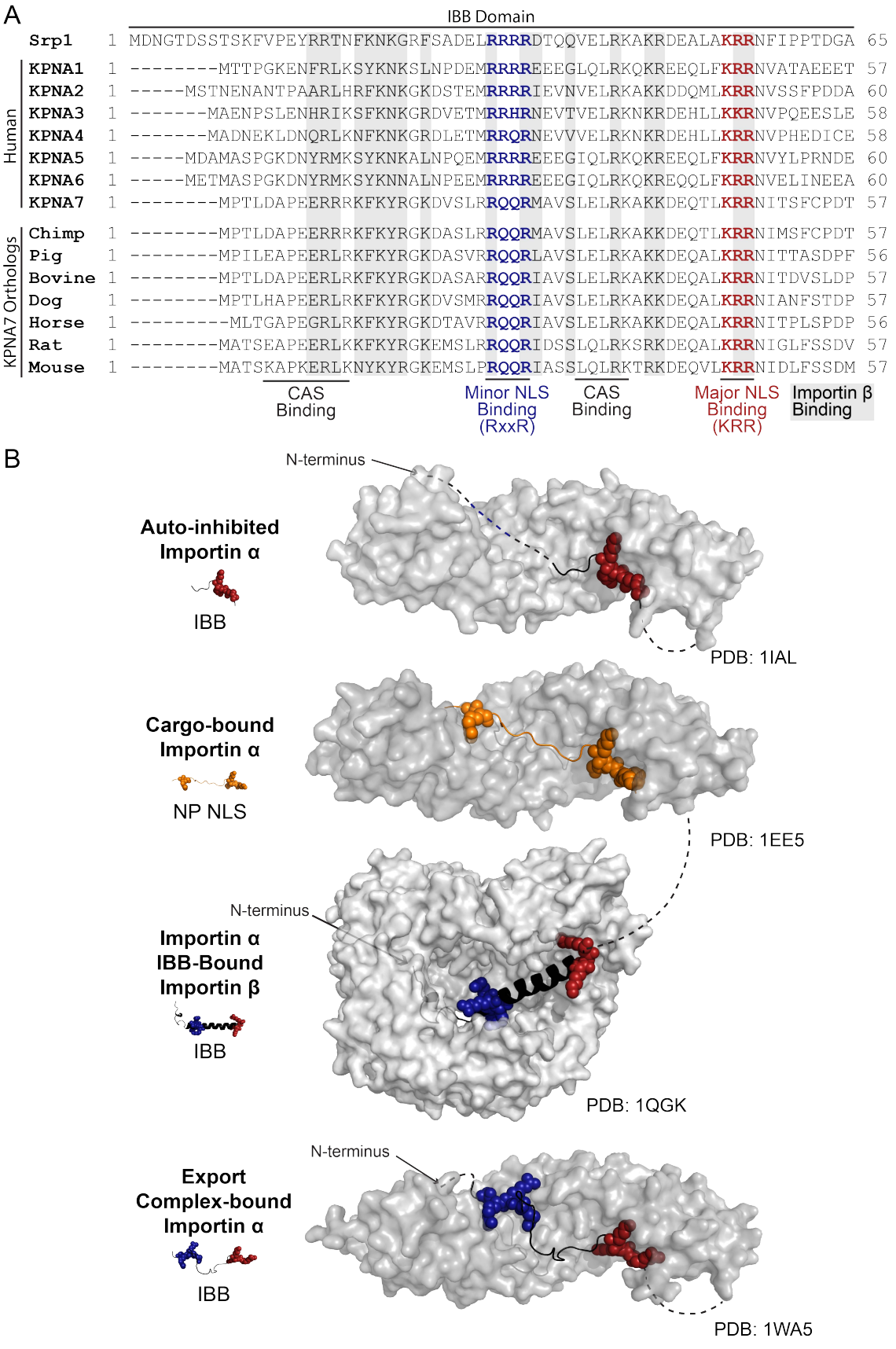

A
IBB Domain
Srp1 1 MDNGTDSSTSKFVPEYRRTNFKNKGRFSADELRRRRDTQQVELRKAKRDEALAKRRNFIPPTDGA 65
Human
KPNA1 1 --------MTTPGKENFRLKSYKNKSLNPDEMRRRREEEGLQLRKQKREEQLFKRRNVATAEEET 57
KPNA2 1 -----MSTNENANTPAARLHRFKNKGKDSTEMRRRRIEVNVELRKAKKDDQMLKRRNVSSFPDDA 60
KPNA3 1 -------MAENPSLENHRIKSFKNKGRDVETMRRHRNEVTVELRKNKRDEHLLKRRNVPQEESLE 58
KPNA4 1 -------MADNEKLDNQRLKNFKNKGRDLETMRRQRNEVVVELRKNKRDEHLLKRRNVPHEDICE 58
KPNA5 1 -----MDAMASPGKDNYRMKSYKNKALNPQEMRRRREEEGIQLRKQKREEQLFKRRNVYLPRNDE 60
KPNA6 1 -----METMASPGKDNYRMKSYKNNALNPEEMRRRREEEGIQLRKQKREQQLFKRRNVELINEEA 60
KPNA7 1 --------MPTLDAPEERRRKFKYRGKDVSLRRQQRMAVSLELRKAKKDEQTLKRRNITSFCPDT 57
KPNA7 Orthologs
Chimp 1 --------MPTLDAPEERRRKFKYRGKDASLRRQQRMAVSLELRKAKKDEQTLKRRNIMSFCPDT 57
Pig 1 --------MPILEAPEERLRKFKYRGKDASVRRQQRLAVSLELRKAKKDEQALKRRNITTASDPF 56
Bovine 1 --------MPTLDAPEERLRKFKYRGKDASARRQQRIAVSLELRKAKKDEQALKRRNITDVSLDP 57
Dog 1 --------MPTLHAPEERLRKFKYRGKDVSMRRQQRIAVSLELRKAKKDEQALKRRNIANFSTDP 57
Horse 1 ---------MLTGAPEGRLRKFKYRGKDTAVRRQQRIAVSLELRKAKKDEQALKRRNITPLSPDP 56
Rat 1 --------MATSEAPEERLKKFKYRGKEMSLRRQQRIDSSLQLRKSRKDEQALKRRNIGLFSSDV 57
Mouse 1 --------MATSKAPKERLKNYKYRGKEMSLPRQQRIASSLQLRKTRKDEQVLKRRNIDLFSSDM 57
CAS Binding
Minor NLS Binding (RxxR)
CAS Binding
Major NLS Binding (KRR)
Importin β Binding
B
N-terminus
Auto-inhibited Importin α
IBB
PDB: 1IAL
Cargo-bound Importin α
NP NLS
PDB: 1EE5
N-terminus
Importin α IBB-Bound Importin β
IBB
PDB: 1QGK
N-terminus
Export Complex-bound Importin α
IBB
PDB: 1WA5

Figure 2.1: Structure and function of the importin-β binding (IBB) domain.
(**A**) Multiple sequence alignment of IBB domains from *S. cerevisiae,* the seven members of the human importin α-family (KPNA1-7), and mammalian orthologs of KPNA7. Sequences known or predicted to mediate auto-inhibition through interactions with the minor NLS binding site (blue) and major NLS binding site (red) are shown in bold. The two regions of the IBB domain known to facilitate binding to the export factor CAS are underlined. The residues which interact with Imp-β are highlighted in grey. (**B**) Structures of full-length mouse Imp-α (PDB 1IAL), yeast Imp-α with the bipartite NPM NLS in the cargo-bound state (PDB 1EE5), (human) Imp-β with the IBB domain (PDB 1QGK), and yeast Imp-α in the export complex-bound stat (PDB 1WA5). The dotted lines depict unstructured regions. The structures, which were solved by other groups (Cingolani et al., 1999b; Conti and Kuriyan, 2000; Kobe, 1999; Matsuura and Stewart, 2004), were rendered using Pymol (Schrodinger LLC, 2015).

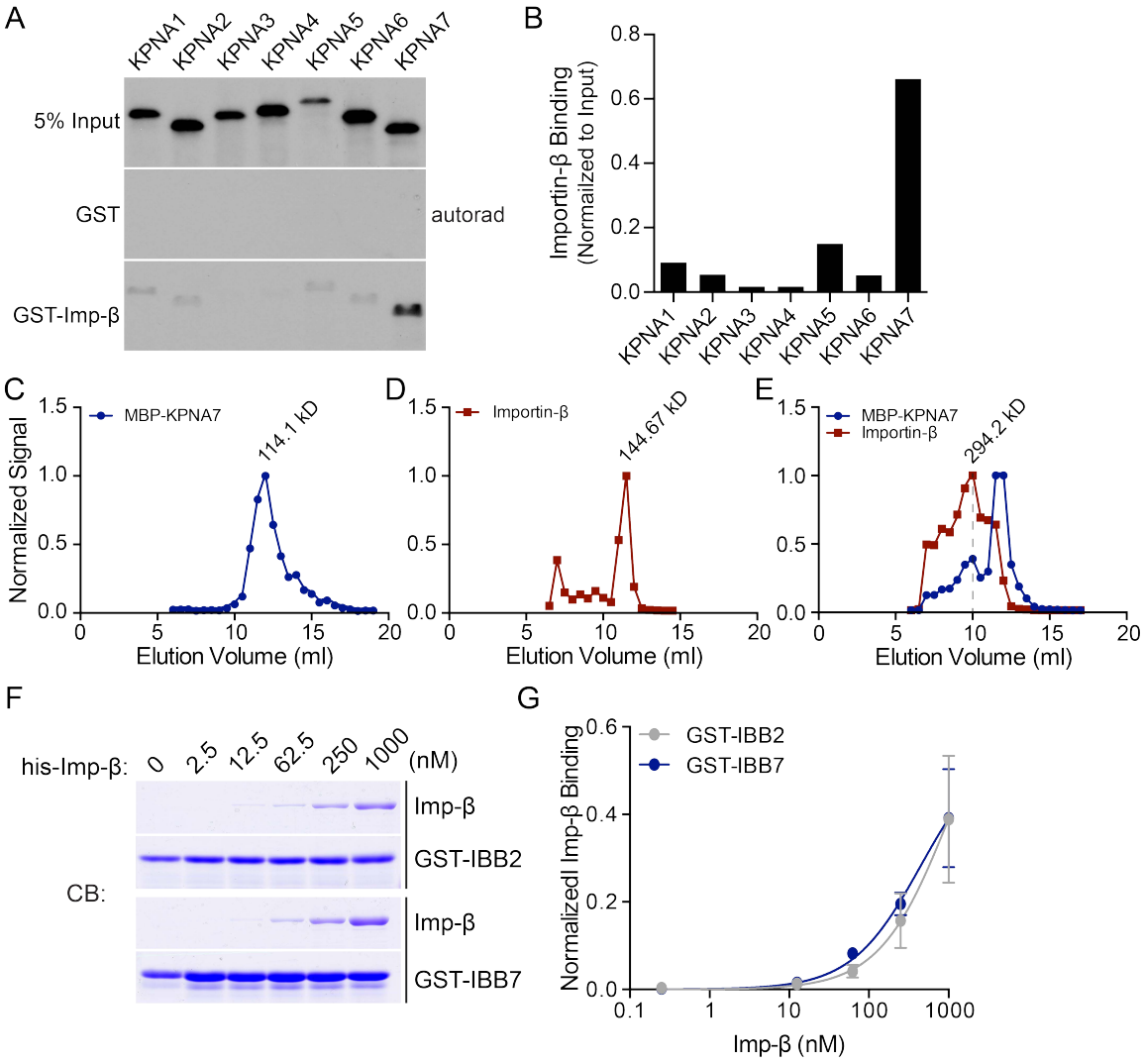


Figure 2.2: KPNA7 shows atypical binding to Imp-β compared to other family members. (**A, B**) Binding assay with ^{35}S-labeled KPNA proteins and GST-Imp-β immobilized on glutathione beads. Bound fractions were analyzed by SDS PAGE and autoradiography, and quantified. (**C-E**) Size exclusion chromatography of KPNA7-Imp-β complexes. MBP-KPNA7 and His-Imp-β were chromatographed individually, and after mixing 1:1. Eluted fractions were analyzed by immunoblotting and the profiles were compared to molecular weight standards (**Fig. S2.1**). (**F**) Isolated IBB domains from KPNA2 and KPNA7 show similar binding to Imp-β. GST-IBB2 and GST-IBB7 were immobilized on glutathione beads and incubated with the concentrations of his-Imp-β indicated in the figure. (**G**) His-Imp-β binding as a function of input concentration was plotted with data averaged from two experiments.

KPNA2, suggesting this chimera has adopted an auto-inhibited state (**Fig. 2.3B, C**). By contrast, KPNA2 containing the KPNA7 IBB maintained the auto-inhibited state of native KPNA2. These data suggest the open-state of KPNA7 is based on sequence and structural information in both the IBB and the ARM repeats. To further characterize the differences in Imp-β binding between KPNA7 and KPNA2 we immobilized GST-KPNA2 and GST-KPNA7 on beads and interrogated binding of his-Imp-β (200 nM) in the presence of increasing amounts of a biotin-BSA-SV40-NLS conjugate (**Fig 2.3D**). The addition of the NLS had minimal effect on KPNA7 binding to Imp-β, while Imp-β binding to KPNA2 was absent without NLS addition and increases proportionally to NLS concentration. This suggests that KPNA7 binding to Imp-β is NLS independent while KPNA2 is not.

C. *The open state of KPNA7 suppresses CAS-mediated export*

Recycling of nuclear Imp-α proteins to the cytoplasm requires Ran-GTP-mediated disassembly of the import complex. RanGTP binding to Imp-β induces its release from Imp-α, dissociation of NLS cargo, and re-establishment of the pre-import, auto-inhibited state. The auto-inhibited conformation of Imp-α is recognized by the export factor CAS, which together with RanGTP assembles into a trimeric complex that undergoes export to the cytoplasm (Matsuura and Stewart, 2004). CAS binds selectively to NLS cargo-free Imp-α by binding to the C-terminal domain of the ARM repeats, as well as regions of the IBB (**see Fig. 2.1A**) that are properly positioned only when the IBB in the auto-inhibited conformation (Matsuura and Stewart, 2004). Our finding that KPNA7 exists largely in the open-state even in the absence of NLS cargo suggests it could show reduced interactions with CAS. We tested this possibility in binding assays and determined that in contrast to KPNA2, KPNA7 was deficient for assembly into a KPNA7-CAS-RanGTP complex (**Fig. 2.4A, B**). This does not reflect an inability of CAS to bind the KPNA7 IBB, since the chimera containing the KPNA7 IBB and KPNA2 ARM repeats showed similar levels of CAS binding as native KPNA2 (**Fig. 2.4C, D**). Thus, partial rescue of CAS binding was observed with the chimera containing the KPNA2 IBB and KPNA7 ARM repeats (**Fig. 2.4C, D**). We conclude that the open-state of KPNA7 is inhibitory to formation of a trimeric complex with CAS and RanGTP.

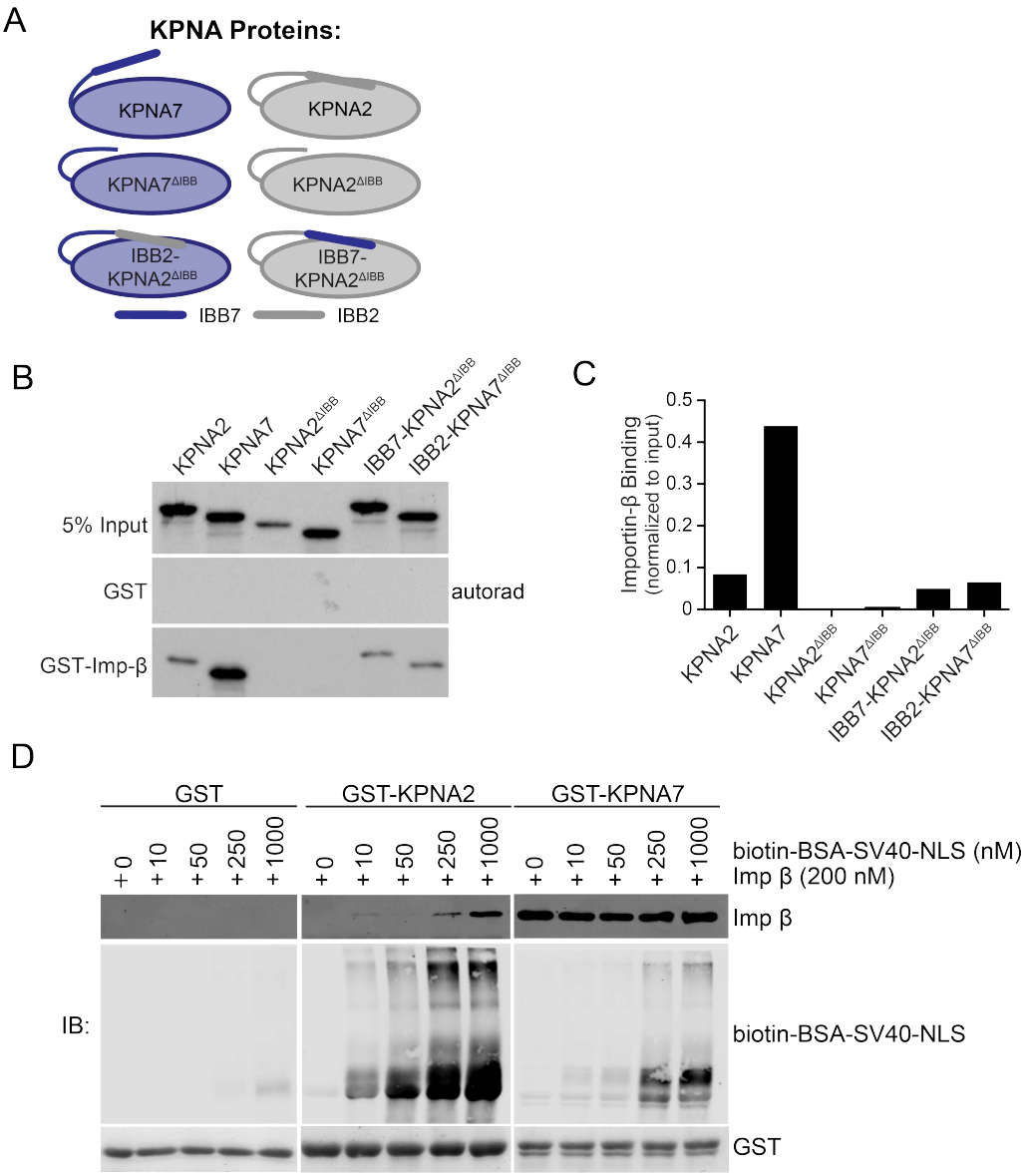


Figure 2.3: Construction and characterization of KPNA-IBB domain chimeras show KPNA7 can adopt a closed state.

(**A**) Cartoons of WT, ΔIBB, and IBB fusions using KPNA2 and KPNA7. The top row depicts the predicted positions of the IBB domain in KPNA7 (open) and KPNA2 (closed). The second row shows the ΔIBB forms of KPNA2 and KPNA7. The third row illustrates closed states of the chimeric KPNA proteins. (**B,C**) Binding assays using ^{35}S-labeled proteins and GST-Imp-β immobilized on glutathione beads. Bound fractions were analyzed by SDS PAGE and autoradiography, and quantified. (**D**) Binding assay of KPNA2 and KPNA7 to Importin-B. GST-KPNA2 and KPNA7 were immobilized on glutathione beads and used to pull-down his-Imp-β (200), Increasing concentrations of a biotin-BSA-SV40 NLS conjugate were used to induce Imp-β binding by KPNA2.

To examine if CAS can mediate nuclear export of KPNA7, which is a critical step in its recycling to the cytoplasm, we used a digitonin-permeabilized cell assay that reconstitutes nuclear transport through the nuclear pore complex. The assay system is widely used for import analysis by measuring nuclear accumulation of fluorescent NLS cargo. In our adaptation, a HEK293T cell line stably expressing HA-KPNA7 was generated, and export was measured as the reduction in nuclear HA-KPNA7 in response to addition of recombinant factors. KPNA2 export was determined in a similar manner by measuring the reduction of the endogenous protein. The assays included recombinant Ran (0.2 μM), the Ran import factor, NTF2 (0.2 μM), an energy regenerating system, and CAS (0-2 μM). After permeabilization and release of the cytosol, and without CAS addition, HA-KPNA7 and KPNA2 were detected by IF microscopy in the nuclei of most cells; quantification of these levels established the baseline for the export assay (**Fig. 2.5A, B**). We found that both HA-KPNA7 and KPNA2 underwent CAS-dependent export, but there was a striking difference in concentrations required for the two receptors. Nuclear export of KPNA2 was virtually complete with the addition of 0.125 uM CAS, whereas 2-10-fold higher concentrations of CAS were required to observe significant KPNA7 export (**Fig. 2.5A, B**). Under these conditions, even the highest concentration of CAS tested (2 uM) which is comparable to estimated levels in cell nuclei (Ribbeck et al., 1998), failed to export all HA-KPNA7 from the nucleus. These data are consistent with the results from biochemical assays and suggest that the open-state of KPNA7 renders it resistant to CAS-mediated export from the nucleus. Furthermore, these data raised the question of whether KPNA7 actively shuttles between the nucleus and cytoplasm, or whether its activity is used in the nucleus and the open-state helps maintain a nuclear distribution. We performed heterokaryon fusion assays (**Fig. 2.6A**) using U2OS cells transfected with Imp-α plasmids and prostate cancer cells expressing the androgen receptor (AR), which is a shuttling transcription factor (Jividen et al., 2018; Ni et al., 2013). By co-staining for AR and the Imp-α receptors, we determined that all nuclei within heterokaryons display positivity for both AR and Imp-α receptors (**Fig. 2.6B, C**). Thus, KPNA7 undergoes recycling to the cytoplasm in intact cells despite the weak interactions with CAS.

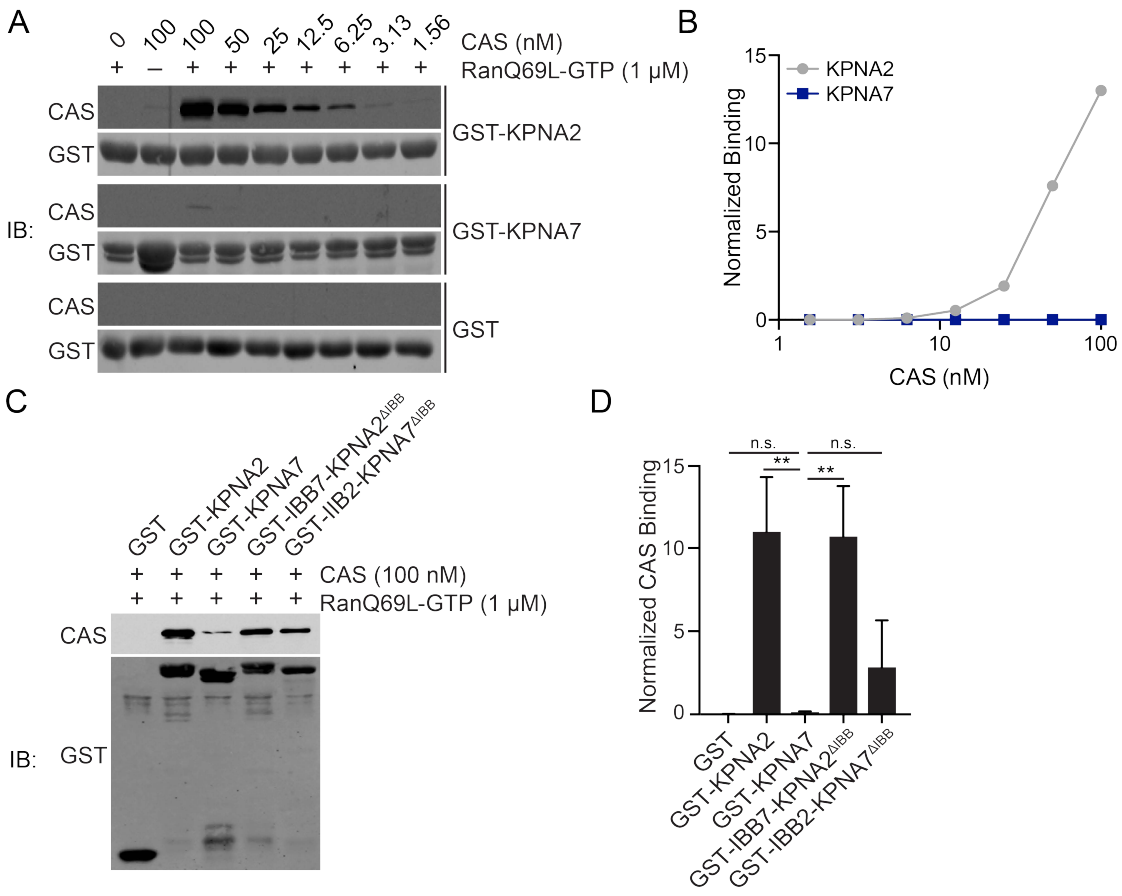


Figure 2.4: KPNA7 displays weak binding to CAS, the export factor for KPNA family members.

(**A, B**) Binding assays using GST-KPNA2 and GST-KPNA7 immobilized on glutathione beads, recombinant CAS, and RanQ69L preloaded with GTP. Bounds fractions were analyzed by SDS PAGE and immunoblotting. The level of CAS binding (normalized to GST) was plotted as a function of CAS input (logarithmic scale). (**C,D**) GST fusions of WT and KPNA-IBB domain chimeras were immobilized on glutathione beads and used for CAS binding assays in the presence of RanGTP. The level of CAS binding determined by immunoblotting was plotted using data averaged from two experiments. ($p<0.01$)

D. *Imp-β enhances KPNA7 binding to NLS cargo*

Auto-inhibition by the IBB greatly reduces NLS binding affinity to KPNA receptors unless Imp-β is present (Harreman et al., 2003a). This helps coordinate transport complex assembly with targeting to, and translocation through the nuclear pore complex-mediated by Imp-β. Our data indicating that KPNA7 maintains an open-state in the absence of NLS cargo raised the important question of whether Imp-β affects KPNA7 import complex assembly. While the key residues which facilitate NLS binding are conserved across all seven Imp-α isoforms, initial studies on KPNA7 indicated it binds very weakly to NLS cargoes recognized by KPNA2 (Kelley et al., 2010). Subsequent discovery efforts by several groups using mass spectrometry identified additional NLS-containing proteins that bind KPNA7 (Kimoto et al., 2015; Park et al., 2012; Vuorinen et al., 2017). Reasoning that some (or possibly all) NLS-cargos might display low-affinity interactions with KPNA7, we established a bead-based fluorescent assay that Rexach and colleagues showed can reconstitute micromolar affinity protein-protein interactions on a bead surface (Patel and Rexach, 2008). In our adaptation of the assay, a maltose-binding protein fusion of KPNA7 (MBP-KPNA7) is immobilized on the surface of maltose beads, combined with GFP-tagged NLS cargoes, and the beads imaged by fluorescence microscopy (**Fig. 2.7A**). We used multiple controls to establish the specificity of the assay. GST-GFP-SV40-NLS, but not GST-GFP, is recruited to the bead surface (**Fig. 2.7B**). To determine if GST-GFP-SV40-NLS binding to the bead surface is KPNA7-dependent, we mixed beads containing KPNA7-MBP with beads containing unfused MBP; a small amount (5%) of IRDye680-labeled MBP was included with the latter to mark the control beads. We observed that GST-GFP-SV40 NLS bound to the KPNA7 beads but not the MBP beads (pink). GST-GFP-SV40 NLS binding to KPNA7 in this assay was enhanced by Imp-β (**Fig. 2.7B-D**). This result was unexpected since the open-form of KPNA7 would be predicted to be fully amenable to NLS binding. Our results from the bead assay suggested that Imp-β might enhance KPNA7 binding to NLS cargo via a mechanism that is distinct from relief of auto-inhibition.

We tested whether Imp-β enhances KPNA7 binding to other NLS cargoes to determine if this property is a feature of binding to both monopartite and bipartite import signals. In addition to the SV40 NLS, we generated recombinant GST-GFP fusions containing the

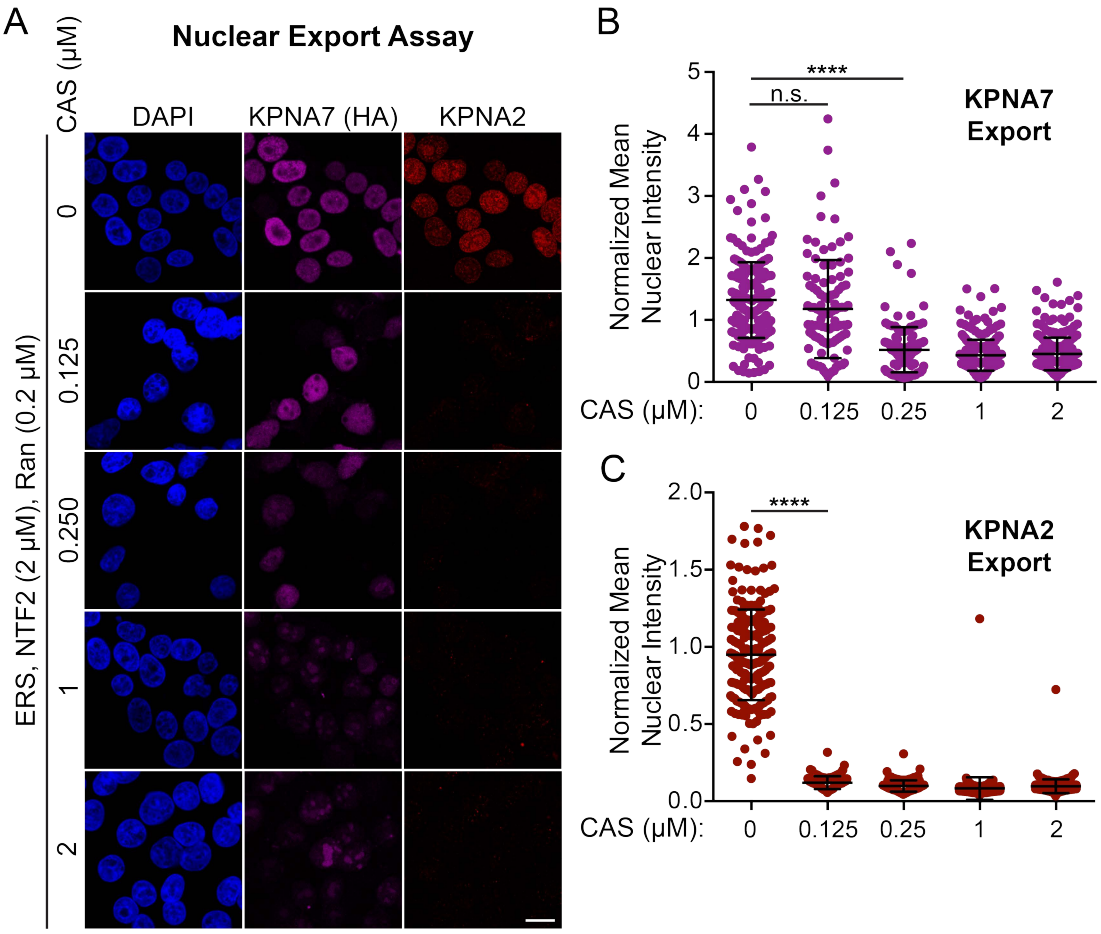


Figure 2.5: Different CAS concentrations are required for nuclear export of KPNA7 and KPNA2.

(**A**) Nuclear export of KPNA7 and KPNA2 assayed in digitonin-permeabilized cells. HEK293T stably expressing HA-KPNA7 were permeabilized with 0.005% digitonin, and incubated with CAS (concentrations indicated), Ran (0.2 μM), NTF2 (2 μM), and an energy regenerating system at 30°C. Post-export reactions were fixed and processed for IF microscopy for KPNA7 and KPNA2, and images were captured by confocal microscopy. (**B, C**) Mean nuclear intensity of KPNA2 and KPNA7 were determined using ImageJ and plotted. (**** $p<0.0001$; Scale: 10 μm)

NLS sequences of Rb, the neural transcription factor Brn2, and the DNA repair protein DDB2 (**Fig. 2.8A, Fig. S2.2A**). We used the fluorescent bead assay, but also standard bead pull-down approach. In the pull-down approach, the SV40 NLS and DDB2 NLS displayed the strongest binding, and both were enhanced by Imp-β addition (**Fig. 2.8B**). Low levels of Rb NLS and Brn2 NLS binding to KPNA7 were detected, the latter enhanced by Imp-β. These data were comparable to results obtained using the fluorescent bead assay, where the strongest binding to KPNA7 was observed with SV40 NLS and DDB2 NLS (**Fig. 2.8C**). In each of these experiments, MBP alone did not display non-specific binding (**Fig. S2.2B-D**). We considered two models to explain how Imp-β enhances KPNA7 interactions with NLS cargo. Although our data is consistent with a constitutively open-state of KPNA7, it is possible (model I) that the IBB contact with the ARM structure in KPNA7 is transient, and Imp-β traps KPNA7 in an open-state. Another possibility (model II) is that Imp-β binding to the IBB of KPNA7 in the open-state induces a structural change that enhances NLS binding. If Imp-β affects cargo binding simply by trapping the open-state of KPNA7 and preventing closure into an auto-inhibited conformation, then removal of the IBB should phenocopy Imp-β addition to the assay. We tested this biochemically by deleting the IBB. The proof-of-principle for this experiment was established by Corbett and colleagues who showed that deleting the IBB from yeast importin-α results in high-affinity binding to NLS that is equivalent to that observed with intact importin-α in the presence of Imp-β (Fanara et al., 2000). We compared GST-GFP-NLS binding to KPNA in the absence and presence of Imp-β, and to Imp-α lacking the IBB, by pull-down assay. We detected small Imp-β enhancement of GST-GFP-NLS binding to KPNA7 over a broad NLS concentration range (90-2400 nM). Deletion of the IBB from KPNA7 dramatically reduced NLS binding, which appears to refute model I and suggests that the IBB is important for the ARM structures involved in binding the NLS (**Fig. 2.8D, E**).

We next tested if Ran-GTP-stimulated release of Imp-β from KPNA7 occurs with an efficiency that is comparable to KPNA2. We generated NLS-KPNA-Imp-β complexes (Assembly Reactions), incubated the complexes with buffer, RanGDP, or RanGTP, and examined Imp-β in the released fraction. Imp-β was released from KPNA7 and KPNA2

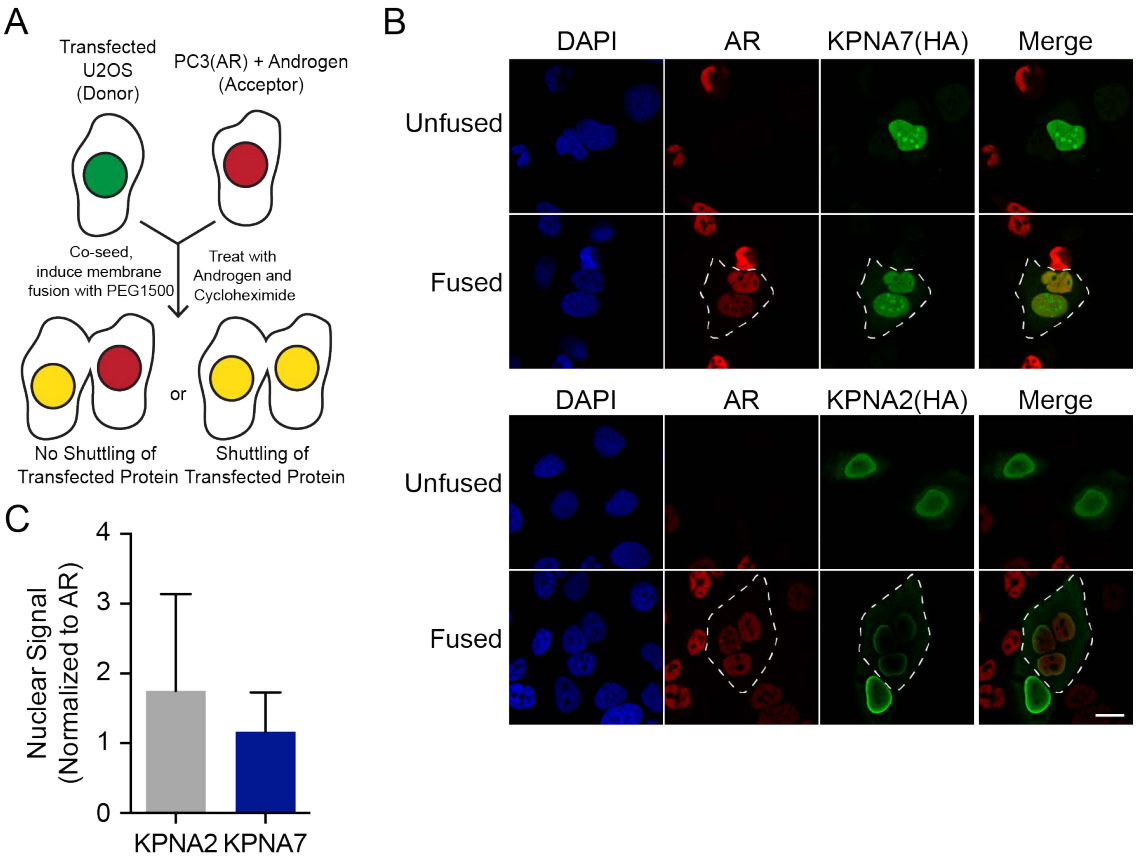


Figure 2.6: KPNA7 undergoes nucleocytoplasmic shuttling in cells.
(**A**) Schematic of cell fusion in heterokaryon assay. HA-KPNA7 and HA-KPNA7 were transfected into U2OS cells. (**B, C**) HA-KPNA2 and HA-KPNA7 transfected into U2OS cells are the donor nuclei in the fusion. Androgen receptor expressing PC3 cells provide the acceptor nuclei. The cells were co-plated and induced to fuse by PEG1500 treatment (Black et al., 2001). Cells were processed for IF and imaged by confocal microscopy. Signal intensity ratios of double-positive nuclei (AR, KPNA) were measured and plotted. (Scale: 10 µm.)

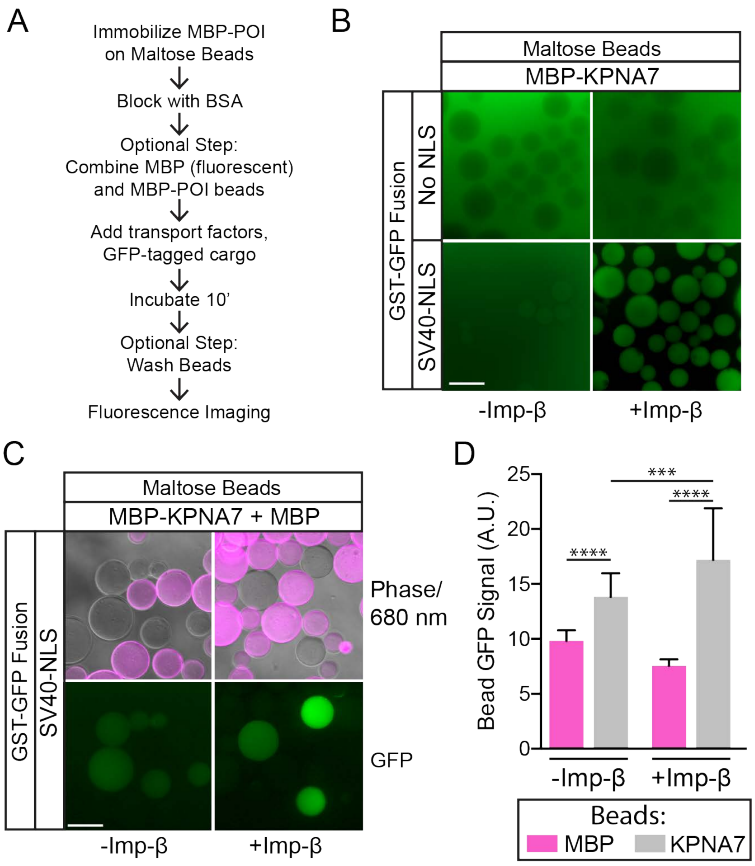


Figure 2.7: Measurement of NLS binding to KPNA7 using bead-based fluorescence imaging.

(**A**) Workflow for bead-based fluorescence binding assay. (**B**) Images from the bead-based fluorescence binding assay using MBP-KPNA7 immobilized on maltose beads and GST-GFP alone and fused to the SV40 NLS (750 nM). Reactions were performed ± Imp-β (1 µM) and were imaged directly without a washing step. (**C**) Bead-based fluorescence binding assay using MBP beads and MBP-KPNA7 beads in the same reaction. MBP beads were loaded with a tracer amount of IRDye680-labeled MBP, which allows identification of the control beads. Beads were mixed at an equal ratio, incubated with GST-GFP-SV40 NLS (750 nM) ± Imp-β (1 µM), washed, and imaged. Mean GFP and 680 nm intensity for each bead was measured. (**D**) Bead-associated fluorescence intensities were measured on MBP beads (680 nm positive) and MBP-KPNA7 beads (680 nm negative), and mean values were plotted. (Scale: 100 µm; n=34, 48, 66, and 16; *** $p<0.001$; **** $p<0.0001$)

in a RanGTP-specific manner (**Fig. 2.8F, G**). This suggests that RanGTP terminates the import reactions of KPNA7 complexes, as it does with other Imp-α complexes.

E. *KPNA7 mediates NLS-dependent nuclear import*

We characterized the import activity of KPNA7 using recombinant proteins and digitonin-permeabilized cell assays. In these assays, we also examined the level of KPNA7 that accumulates in the nucleus by IF microscopy. We found that KPNA7 mediated nuclear import of GST-GFP-SV40 NLS (**Fig. 2.9A**). The effect was specific since the NLS was required for the cargo to accumulate in the nucleus, and relatively little GST-GFP-SV40 NLS entered the nucleus without KPNA7 and Imp-β addition (**Fig. 2.9A**). We also tested whether nuclear import of a cargo that binds weakly to KPNA7 can be observed in this assay (Li et al., 2015; Park et al., 2012). The cargo protein GST-GFP-Brn2 NLS underwent nuclear import in a manner that was dependent on a functional NLS (**Fig. 2.9B, C**) since import was reduced significantly by a single amino substitution in the NLS (K356E).

F. *Cargo-free KPNA7 can enter the nucleus*

Because KPNA7 can bind Imp-β in the absence of NLS cargo, our data raises the question of whether KPNA7 can enter the nucleus in a cargo-free state. This view was implied by KPNA7 entry into nuclei of permeabilized cells which occurred without adding NLS cargo. A single amino acid substitution in yeast importin-α (D203K) was shown to reduce NLS binding affinity by 300-fold (Leung et al., 2003). By protein modeling using a crystal structure of KPNA2 (PDB:1EJY) and a threaded structure of KPNA7 (modeled on PDB:1IAL with SWISS-MODEL), we identified the residues in the mammalian proteins that are likely equivalent to D203 in yeast importin-α (**Fig. 2.10A, B**). We generated the single amino acid substitutions in KPNA7 (D185K) and KPNA2 (D192K) and analyzed the distribution of the epitope-tagged proteins by IF confocal scanning microscopy. We determined that the NLS binding mutant form of KPNA7, as well as KPNA2, both entered the nucleus (**Fig. 2.10C**). This suggests that Imp-α proteins may enter the nucleus in a cargo-free form. The D192K mutation in KPNA2 shifted its nuclear/cytoplasmic ratio, likely because of a reduced ability to adopt the auto-inhibited conformation required for nuclear export by CAS. Introduction of the same amino acid

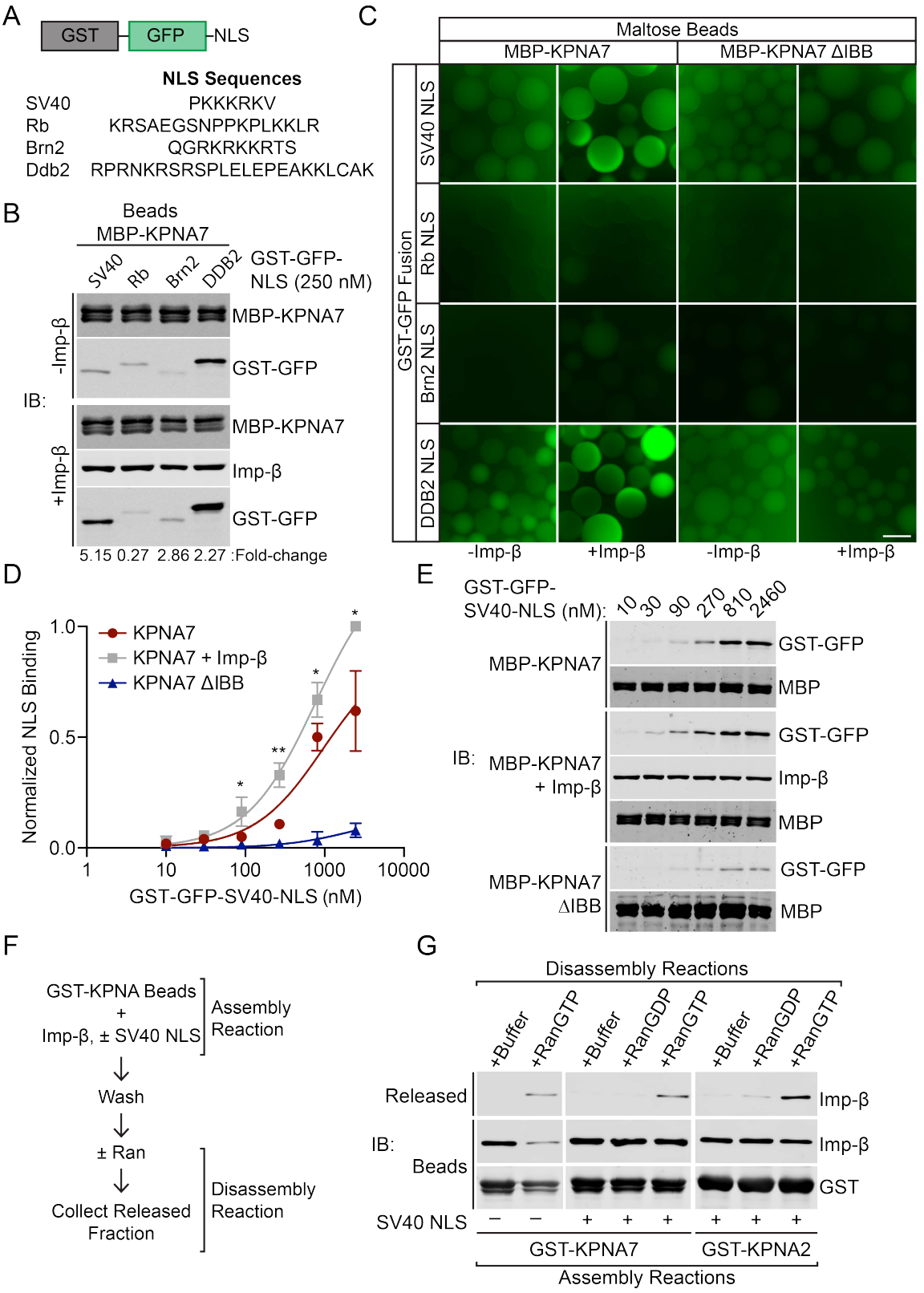

A
GST
GFP
NLS
NLS Sequences
SV40 PKKKRKV
Rb KRSAEGSNPPKPLKKLR
Brn2 QGRKRKKRTS
Ddb2 RPRNKRSRSPLELEPEAKKLCAK
B
Beads
MBP-KPNA7
SV40
Rb
Brn2
DDB2
GST-GFP-NLS (250 nM)
-Imp-β
+Imp-β
IB:
MBP-KPNA7
GST-GFP
MBP-KPNA7
Imp-β
GST-GFP
5.15 0.27 2.86 2.27 :Fold-change
C
Maltose Beads
MBP-KPNA7
MBP-KPNA7 ΔIBB
GST-GFP Fusion
SV40 NLS
Rb NLS
Brn2 NLS
DDB2 NLS
-Imp-β
+Imp-β
-Imp-β
+Imp-β
D
Normalized NLS Binding
KPNA7
KPNA7 + Imp-β
KPNA7 ΔIBB
1.0
0.5
0.0
1
10
100
1000
10000
GST-GFP-SV40-NLS (nM)
E
GST-GFP-SV40-NLS (nM):
10
30
90
270
810
2460
MBP-KPNA7
GST-GFP
MBP
IB:
MBP-KPNA7 + Imp-β
GST-GFP
Imp-β
MBP
MBP-KPNA7 ΔIBB
GST-GFP
MBP
F
GST-KPNA Beads + Imp-β, ± SV40 NLS
Assembly Reaction
Wash
± Ran
Collect Released Fraction
Disassembly Reaction
G
Disassembly Reactions
+Buffer
+RanGTP
+Buffer
+RanGDP
+RanGTP
+Buffer
+RanGDP
+RanGTP
Released
Imp-β
IB:
Beads
Imp-β
GST
SV40 NLS
– – + + + + + +
GST-KPNA7
GST-KPNA2
Assembly Reactions

Figure 2.8: KPNA7 binding to NLS cargoes is enhanced by Imp-β.
(**A**) Diagrams of GST-GFP fusions engineered with NLS sequences from four different proteins. GST-GFP-NLS proteins were purified and examined by SDS PAGE (**Fig. S2.2A**). (**B**) Pull-down assays with MBP-KPNA7 (see **Fig. S2.2B** for MBP) immobilized on maltose beads. The assays, which contained GST-GFP-NLS cargoes and Imp-β (500 nM), were analyzed by immunoblotting. GST signal was normalized to MBP and the fold-change in binding after Imp-β addition is shown. (**C**) Bead-based fluorescence binding assay using immobilized MBP-KPNA7 and MBP-KPNA7 ΔIBB. Beads were incubated with GST-GFP-NLS fusions (750 nM) with and ± Imp-β (1 μM), washed once, and imaged. The complete set of reactions including phase contrast images and MBP control beads is provided (**Fig. S2.2D**). Scale bar 100 μm. (**D, E**) Pull-down assay showing GST-GFP-SV40 NLS binding to KPNA7 ±Imp-β, and to KPNA7 ΔIBB. GST signal was normalized to MBP signal from three independent experiments (one replicate in (**E**)). Significance is indicated between KPNA7 and KPNA7 plus Impβ. (* $p<0.05$, ** $p<0.01$) (**F**) Scheme for testing RanGTP dissociation Imp-β from KPNA7 and KPNA2. (**G**) Dissociation of Imp-β from KPNA7 and KPNA2 using RanGTP analyzed by immunoblotting.

substitutions into the KPNA7-KPNA2 chimeras generated the same effect seen in KPNA2, an increase in the nuclear/cytoplasmic ratio (**Fig. 2.10C, D**). The increased nuclear concentration is predicted to result from reduced CAS interactions with mutants that fail to adopt an auto-inhibited, closed state.

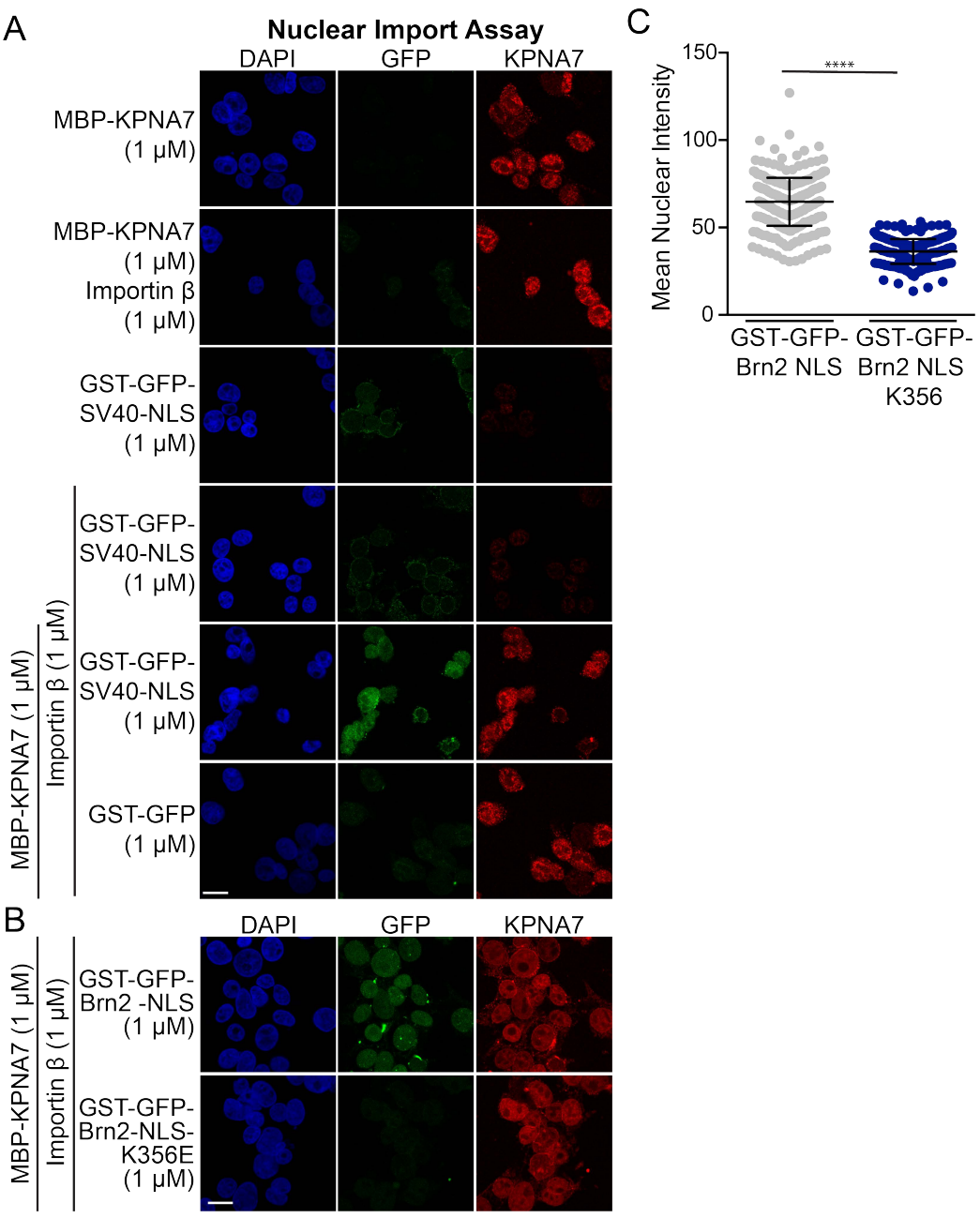


Figure 2.9: KPNA7 mediates nuclear import by selective recognition of a functional NLS. (**A**) Fluorescence microscopy images of permeabilized cell import assays showing nuclear accumulation of GST-GFP-SV40 NLS (GFP) and recombinant MBP-KPNA7 (red). The assays were performed in HEK293T permeabilized with 0.005% digitonin and incubated with Ran (0.2 μM), NTF2 (2 μM), an energy regenerating system, and recombinant factors as indicated in the panel. Post-import reactions were fixed and processed for IF microscopy for KPNA7, and images were captured by confocal microscopy. (B) Fluorescence microscopy images of permeabilized cell import assays using GST-GFP-Brn2-NLS (WT and mutant, K356E), performed and imaged as in panel (A). (**C**) Quantification of KPNA7-mediated nuclear import of Brn2 NLS (n=310 nuclei) and Brn2 NLS (K356E) (n=232 nuclei) by KPNA7. Scale bar: 10 μm; **** $p<0.0001$).

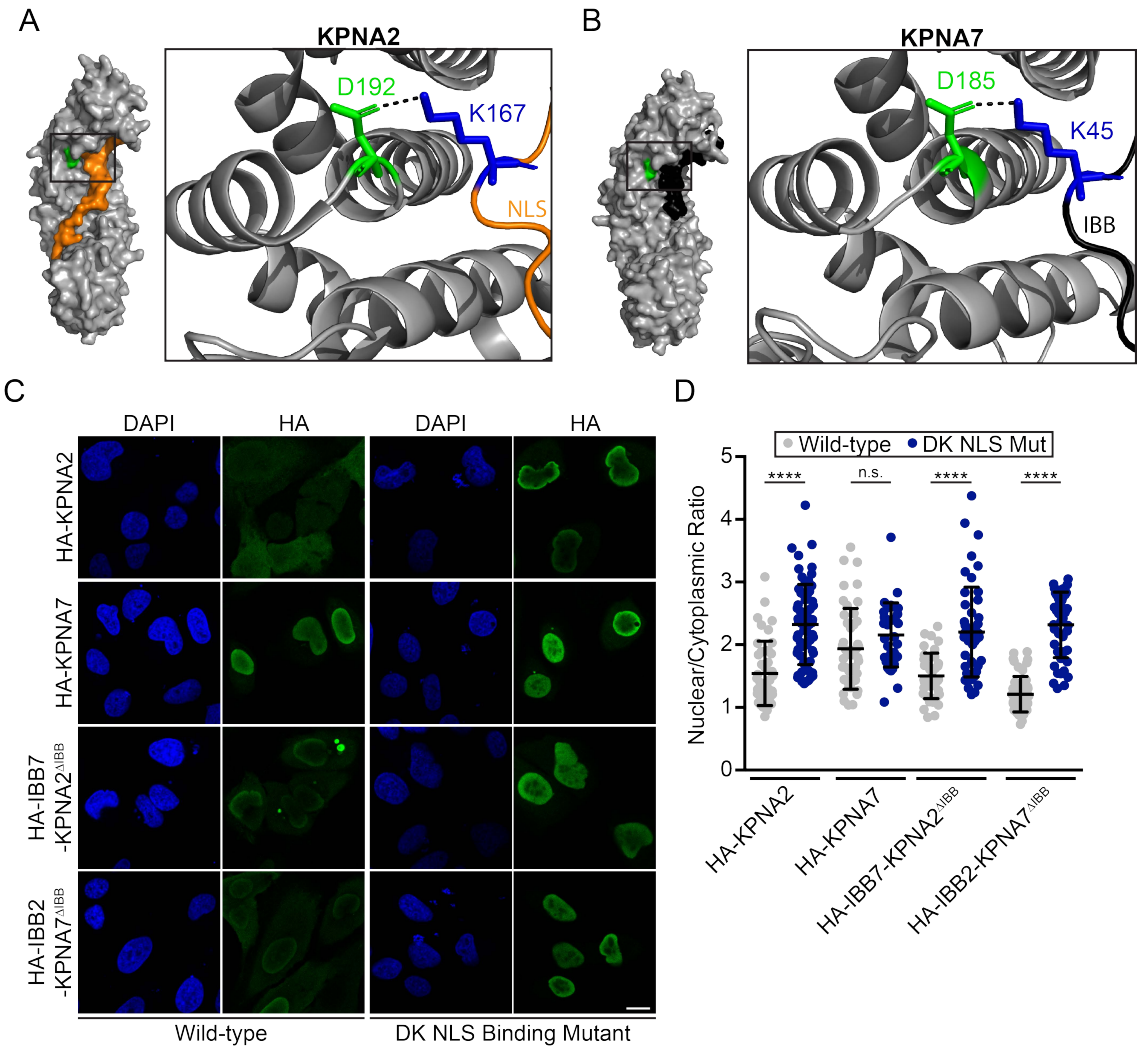


Figure 2.10: KPNA7 and KPNA2 can undergo nuclear import independent of NLS binding. (**A**) Structural model (surface view) of mouse Kpna2 bound to the Rb NLS (orange) (PDB: 1EJY). Kpna2 Asp192 (green) forms a hydrogen bond with Rb NLS Lys167 (blue). (**B**) Structural model (surface view) of KPNA7 threaded onto mouse Kpna2 (PDB: 1IAL) using SWISS-MODEL. The IBB domain (black) illustrates the auto-inhibited state and shows a predicted hydrogen bond between KPNA7 Asp185 (green) and IBB Lys45 (blue). (**C**) IF microscopy showing the localization of KPNA isoforms and KPNA chimeras with WT and Asp to Lys substitutions ARM4 that disrupt both NLS and IBB binding. The constructs were transfected into U2OS cells, stained via the HA epitope, and imaged by confocal microscopy. (**D**) Plots of nuclear to cytoplasmic ratios (HA signal) for each of the constructs. (n=34-63 nuclei per construct). (**** $p<0.001$)

IV. Discussion

Auto-inhibition of Imp-α by the IBB domain, and relief through Imp-β binding to the IBB domain, is an evolutionarily conserved mechanism. Mutant alleles in the IBB that selectively relieve auto-inhibition are inviable in *S. cerevisiae* (Harreman et al., 2003b, 2003a). This presumably reflects the fact that once the import complex has entered the nucleus, efficient release of NLS cargo requires a specific contribution from the IBB domain (Harreman et al., 2003a). Additionally, CAS-mediated export of Imp-α requires a closed, auto-inhibited state of the receptor (Matsuura and Stewart, 2004).

The crystal structure of full-length mouse Kpna2 indicated that residues 44-54 of the IBB bound in the major NLS binding groove to facilitate an auto-inhibited state, while residues 1-43 and 54-70 could not be modeled, indicating they may be in a disordered state (Kobe, 1999). This structures, together with mutational analysis of the IBB domain, identified the residues that mediate auto-inhibition, mainly the KRR motif, through interactions with the major NLS groove (Harreman et al., 2003a; Kobe, 1999) (**Fig. 2.1A**). Additionally, the structure of the Imp-α export complex with Imp-α bound to CAS and RanGTP demonstrated that residues 33-36, the RxxR motif, bind in the minor NLS binding groove (Matsuura and Stewart, 2004). The KRR motif is perfectly conserved in human KPNA proteins, but there is variation in the RxxR motif. While most KPNA proteins contain four arginine residues within the RxxR motif, KPNA3 contains a histidine residue and KPNA3 and KPNA7 contain one and two glutamines at the variable positions respectively. These residues were among those disordered in the crystal structure of full-length Imp-α (Kobe, 1999), which suggests they do not stably bind the NLS binding groove in a closed Imp-α but does not rule out an impact on auto-inhibition. Thus, the primary sequence of IBB domains raises the interesting question of whether there might be variations in the extent of auto-inhibition of different Imp-α isoforms.

Recent crystallographic work from the Cingolani group examining the IBB domain and auto-inhibition of KPNA4 and KPNA6 identified key features in the ARM repeats that are important for auto-inhibition (Pumroy et al., 2015). The authors of this study identified specific helix-breaking glycine residues in eight of the ten ARM repeats of KPNA2 which play an important role in determining the curvature and flexibility of the

ARM repeat core. Substitution of this glycine for a polar or charged residue, which occurs in multiple ARM repeats in KPNA4 and KPNA6, results in receptors with a higher degree of flexibility in the ARM repeat core of these receptors, as compared to KPNA2. In combination with the sequence differences in the IBB domain, this flexibility allows the viral bipartite NLS from Influenza A virus polymerase subunit PB2, which contains an independently folded globular domain proximal to the NLS, to overcome auto-inhibition and bind KPNA4 in the absence of Imp-β (Pumroy et al., 2015; Tarendeau et al., 2008). This was proposed to be an adaptation of the viral polymerase to promote rapid nuclear translocation by simplifying the transport reaction from trimolecular to pseudo-bimolecular (Pumroy et al., 2015). It should be noted, however, that KPNA4 clearly adopts an auto-inhibited state because binding to the bipartite NLS from nucleoplasmin requires Imp-α.

Our data on KPNA7 suggests it occupies a constitutively open-state, and that it is not governed by auto-inhibition. In contrast to other human Imp-α proteins, KPNA7 displays strong NLS cargo-independent Imp-β binding. Sequence analysis of the KPNA7 IBB and ARM repeat core reveals sequence features that potentially contribute to the open-state. In addition to the variation in the KPNA7 RxxR motif (described above), the KPNA7 ARM repeat core contains two non-glycine helix-breaking residues in ARMs four and eight (**Fig. S2.3).** Following the logic articulated for KPNA4, there are probably features in both the IBB domain and the ARM repeats of KPNA7 that underpin its open state. Consistent with this view, IBB domain swapping experiments showed that the KPNA7 IBB domain can close KPNA2, and the KPNA2 IBB domain can only partially close KPNA7.

In addition to human KPNA4 (Pumroy et al., 2015), there are other examples of Imp-α isoforms where the extent of auto-inhibition might be modulated for functional reasons. These include the apicomplexan parasites *P. falciparum* and *T. gondii,* and plant *A. thaliana* (Bhatti and Sullivan, 2005; Dey and Patankar, 2018; Hübner et al., 1999; Pumroy et al., 2015). Comparison of the IBB domain of each of these Imp-α homologs indicates differences in both the RxxR motif and the KRR motif (**Fig. S2.4B**). The single Imp-α in the parasite *P. falciparum* features the sequence SKR at the position of the KRR

motif. This sequence is necessary to prevent auto-inhibition of the *P. falciparum* Imp-α (Dey and Patankar, 2018). While the relative importance of the open state for Imp-α function in parasites has not been explicitly defined, in *A. thaliana* it was shown that the open state permits the receptor to facilitate nuclear import of NLS cargo independent of Imp-β (Hübner et al., 1999).

Within a species, why does a single Imp-α receptor (or subset) have a predominantly open state, and how does this contribute to import function? It seems likely that the answers to these questions involve the cargoes recognized and carried by these receptors. It is abundantly clear that cargo selectivity, though poorly understood, is an important aspect of nuclear transport regulation. Indeed, expression changes of Imp-α proteins are known to help drive multiple developmental programs based on differential transport of NLS containing proteins (Goldfarb et al., 2004; Köhler et al., 1997; Mason et al., 2002; Pumroy and Cingolani, 2015; Yasuhara et al., 2006). KPNA7, for example, is expressed at low levels in most cell types, but it is highly expressed in the oocyte and during early embryogenesis. In multiple species, KPNA7 has been shown to be required for proper embryogenesis and progression to blastocyst (Tejomurtula et al., 2009; Wang et al., 2012). This suggests that KPNA7 is specifically required to import factors that play critical roles in the developmental program. We speculate the open state of KPNA7 might be important for accommodating the binding and import of proteins with relatively weak import signals, or that pre-formed KPNA7-Imp-β complexes are primed for transport and this increases transport efficiency. This may be broadly similar to how reduced auto-inhibition benefits viral and parasitic protein import (Bhatti and Sullivan, 2005; Dey and Patankar, 2018; Pumroy et al., 2015).

Our data showing that the KPNA7 open state makes it resistant to CAS-mediated export argues it maintains an open state inside the nucleus. It remains possible that KPNA7 adopts a closed state in vivo through an unknown mechanism which our assay conditions fail to recapitulate. Nonetheless, our data raise the interesting possibility that KPNA7 could have functions within the nucleus that involve engaging with NLS-containing proteins. There is a precedent for Imp-α proteins having nuclear functions. In *Xenopus* eggs, Imp-α was shown to bind LaminB, and exogenous Imp-α inhibited lamina assembly

during interphase egg nuclear assembly assays (Adam et al., 2008). This interaction is dependent on the LaminB NLS and is sensitive to RanGTP. During apoptosis, cleavage of the IBB domain of KPNA2 can result in a ΔIBB form of the protein that binds the MCM replication licensing factor to repress DNA synthesis (Kim and Lee, 2008). Specifically for KPNA7, MS/MS data on KPNA7 interacting proteins showed enrichment in proteins with functions in RNA processing and metabolism, and ribonucleoproteins (Kimoto et al., 2015; Vuorinen et al., 2017). While these could be import substrates for KPNA7, they could also be proteins regulated by KPNA7 binding inside the nucleus. Additionally, knockdown of KPNA7 in pancreatic cancer cells results in changes in nuclear morphology, and reorganization of the lamina suggesting the increased levels of KPNA7 in those cell lines helps maintain the integrity of the nuclear envelope (Vuorinen et al., 2018). An alternative function for the KPNA7 open-state was recently suggested by the Yoneda group. Their study identified the ability of recombinant KPNA7 to bind other Imp-α proteins and proposed these heterodimers function to suppress NLS binding due to KPNA7 occupying the NLS binding site (Kimoto et al., 2015; Miyamoto and Oka, 2016). We did not detect evidence of KPNA7 homodimer formation by gel filtration chromatography (**Fig. 2.2C**). Regardless, Imp-α heterodimer formation potentially represents another mechanism for regulation NLS binding that might be used by KPNA7.

In conclusion, KPNA7 adopts an open-state that permits constitutive binding to Imp-β. This suggests that in contrast to most other Imp-α isoforms, the IBB domain of KPNA7 does not lock the receptor in an auto-inhibited state. While KPNA7 can function as a nuclear transport receptor, we speculate that the open state might be a clue that it also has functions within the nucleus.

V. Materials and Methods

A. *Plasmid Construction*

Human KPNA7 was codon-optimized (Genewiz) for *E. coli*, cloned into pMBPHis-Parallel1 (Sheffield et al., 1999), and expressed as a maltose binding protein (MBP) fusion. KPNA7 ΔIBB (a.a. 58-517) was engineered in the same manner. The IBB domains from human KPNA2 (a.a.1-60; IBB2) and human KPNA7 (a.a.1-57; IBB7)

were cloned into pGEX-4T1 (GE Healthcare) for expression as GST fusions. For expression of KPNA chimeric proteins, the ARM repeat domains of KPNA2 (a.a. 61-529) and KPNA7 (a.a. 58-517) were cloned into pGEX-4T1 to yield pGEX-IBB2-KPNA7$^{\Delta IBB}$ and pGEX-IBB7-KPNA2$^{\Delta IBB}$. For in vitro translation and mammalian expression, the chimeric constructs were cloned into pCMVTNT with HA tags, to generate HA-IBB2-KPNA7$^{\Delta IBB}$ and HA-IBB7-KPNA2$^{\Delta IBB}$. An Asp to Lys substitution at position 203 of yeast importin-α has been shown to decrease NLS binding by greater than 300-fold (Leung et al., 2003). Using protein modeling, we designed and introduced the equivalent substitutions in KPNA2 and KPNA7 by site-directed mutagenesis. This yielded HA-KPNA2(D192K), HA-KPNA7(D185K) and HA-IBB2-KPNA7$^{\Delta IBB}$ (D185K) and HA-IBB7-KPNA2$^{\Delta IBB}$ (D192K) plasmids. HA-tagged human KPNA isoforms in pCMVTNT (Promega) have been described (Kelley et al., 2010). Plasmids for expression of recombinant Imp-β, Ran, RanQ69L, NTF2, and CAS have also been published (Chatterjee and Paschal, 2015; Kelley et al., 2010; Kutay et al., 1997). For lentiviral transduction, HA-KPNA7 was cloned into the pWPI (gift from Didier Trono; Addgene #12254) vector using SwaI and PacI restriction sites.

B. *Protein Expression*

MBP fusion proteins were expressed in BL21(DE3) *E. coli* and purified by standard methods. Cultures were grown in LB media overnight at 18°C, harvested, lysed by French press, clarified, and bound to amylose resin (NEB). MBP-tagged proteins were eluted with 10 mM maltose and dialyzed into 50 mM Tris pH 8.0, 200 mM NaCl with 1 mM DTT. GST fusion proteins were expressed in BL21 *E. coli* and purified using glutathione resin (Millipore Sigma) by standard methods. GST tagged proteins were prepared in a manner similar to MBP fusions, using glutathione resin and elution with 10 mM glutathione. All proteins were dialyzed into PBS with 1 mM DTT, examined by SDS-PAGE, and quantified by Bradford protein assay. Methods for expression and purification of Imp-β, Ran, NTF2, and CAS have been published (Chatterjee and Paschal, 2015; Kelley et al., 2010; Kutay et al., 1997).

C. *Binding Assays*

GST and GST-KPNA fusions and derivatives were pre-bound for at least 3 hours at 4°C to glutathione resin at 1 µg/µl bead volume and blocked for 1 hour with 1 mg/ml BSA. Binding assays (and dissociation assays) were performed with recombinant Imp-β, CAS, and Ran using concentrations shown in the figures with 10 µl of loaded beads and 200 µl total reaction volumes. Binding reactions were performed by incubating beads with recombinant proteins for 2 hours at 4°C in PBS, 1 mM DTT, 0.05% Triton-X 100 with leupeptin, pepstatin, and aprotinin (1 µg/ml each). The beads were recovered by brief centrifugation, washed 5X in the same buffer, and analyzed by SDS-PAGE and either Coomassie blue staining or immunoblotting. Quantification for all blots was performed with ImageStudio software (LiCor). In CAS binding and Ran dissociation assays, the GTPase deficient RanQ69L mutant was used, pre-loaded with GTP. For binding assays involving MBP fusions, MBP, MBP-KPNA7 and MBP-KPNA7 ΔIBB were pre-bound to amylose resin for 3 hours at 4°C at a concentration 1 µg/µl packed bead volume and blocked for 1 hour by incubation with 1 mg/ml BSA. Binding reactions were performed in 20 mM Tris-HCl (pH 7.4), 0.2 M NaCl, 1 mM EDTA, 1 mM DTT, 0.05% Triton-X 100 with leupeptin, pepstatin, and aprotinin (1 µg/ml each). Binding reactions, washing, and analysis were the same as described for glutathione bead-based assays.

For radiolabeled binding assays, ^{35}S-KPNA proteins were generated by *in vitro* translation in the presence of ^{35}S-methionine (TnT Coupled Reticulocyte Lysate System; Promega). Labeling efficiency was determined by scintillation counting. KPNA protein binding to Imp-α was assayed using GST-Imp-β immobilized on glutathione resin (1 µg/µl of packed resin as above). Reactions (100 µl total) contained 10 µl of beads and 125,000 cpm of each labeled KPNA protein and were performed in 25 mM Tris pH 7.5, 50 mM NaCl, 5 mM MgCl2, 0.1 mM EDTA, 0.5 mg/ml BSA, 0.1% NP-40, and 1 mM DTT with leupeptin, pepstatin, and aprotinin (1 µg/ml each) for 3 hours at 4°C. Beads were washed 5X in the same buffer, analyzed by SDS-PAGE, and developed on film. Quantification was performed with ImageJ software.

D. *Protein modeling and sequence alignments*

The crystal structures for yeast Importin-α in complex with the nucleoplasmin NLS (PDB: 1EE5) and CAS (PDB: 1WA5), human Importin-β in complex with the KPNA

IBB domain (PDB: 1QGK), and mouse Kpna2 in complex with the Retinoblastoma NLS (PDB: 1EJY) were rendered using PyMol (Schrodinger LLC, 2015). To generate the KPNA7 model structure with IBB bound, the KPNA7 primary amino acid sequence was threaded onto the crystal structure of mouse Kpna2 in complex with the IBB (PDB: 1IAL) using SWISS-MODEL (Waterhouse et al., 2018). Multiple sequence alignments for the human KPNA proteins and select mammalian orthologs of KPNA7 were generated using Clustal-Omega (Chojnacki et al., 2017).

E. *Gel filtration chromatography*

Gel filtration analysis of MBP-KPNA7 and his-Imp-β was performed using an AKTA FPLC System with a Superdex200 10/300 GL column (GE Healthcare) in 50 mM $NaPO_4$ pH 7.0, 150 mM NaCl with 1 mM DTT at a flow rate of 0.25 ml/min. Prior to chromatography, MBP-KPNA7 (100 µg) and his-Imp-β (100 µg) were clarified in an air-driven ultracentrifuge (Beckman Coulter) before injecting onto the Superdex200 column. For complex analysis, MBP-KPNA7 (100 µg) and his-Imp-β (100 µg) were combined and incubated at 4°C for 1 hour prior to clarification. Fractions (0.25 ml) were collected and subjected to SDS-PAGE and immunoblotting. Blots were quantified using ImageStudio (Li-Cor). Molecular weight standards (Sigma) were analyzed in a similar manner and used to generate a standard curve (GraphPad Prism) (**Fig S2.1B**).

F. *Bead-based fluorescence binding assay*

Bead-based fluorescence binding assays were adapted from published methods (Patel and Rexach, 2008). Amylose resin was pre-loaded with protein (1 µg/µl packed beads) for at least 4 hours at 4°C in 20 mM Tris pH 8, 200 mM NaCl, 1 mM EDTA, 1 mM DTT with leupeptin, pepstatin, and aprotinin (1 µg/ml each) and blocked with 5 mg/ml BSA. Beads were resuspended to a 50% slurry. Each binding assay was assembled as a droplet on glass slides by dispensing the following: 1.5 µl bead slurry, 1 µl 4X reaction buffer (40 mM EDTA, 40 mg/ml BSA, 500 mM NaCl), x µl GST-GFP cargo (750 nM, final concentration), x µl Imp-β (750 nM, final concentration), and distilled water to a total volume of 4 µl. A cover glass was placed on the droplet, the slide was incubated for 5 minutes at room temperature, and imaging was performed with EVOS Fl Imaging System

(Invitrogen) with 20x objective. All images for each experiment were captured at equal intensity and time of exposure, chosen to maximize signal without saturation.
A modification of the bead-based fluorescence binding assay was used that included coating the control beads with fluorescently-labeled MBP protein. By this approach, beads containing the protein of interest and the control beads (MBP) can be added to the same reaction; the control beads are distinguished by fluorescent MBP. For this modification of the assay, MBP was labeled with IRDye 680 fluor according to the manufacturer's instructions (Li-Cor). MBP-680 was added to control beads at 0.05 ug/µl during bead loading. MBP-KPNA7 and MBP-680 beads were mixed at a 1:1 ratio and reactions were assembled in the same manner as described above, except in microfuge tubes. Reactions were transferred to 96 well plates and washed 2x for 5 min with 200 µl of 1x reaction buffer and spun at 100x g for 2 min after each wash to bring beads to the bottom of the well. Supernatant was carefully removed by pipetting. After final wash, ~50 µl of buffer was left in the well. Beads were then imaged with phase contrast and GFP and Cy5 filter on EVOS Fl Imaging System (Invitrogen) with 20x objective at equivalent intensity and exposure time. The GFP and 680 intensity of each bead was quantified using ImageJ software. Beads were separated into 680 positive and negative (MBP and MBP-KPNA7 beads respectively) and GFP signal was plotted.

G. *Cell culture*

HEK293T cells (ATCC) were cultured in DMEM (Gibco) with 5% FBS (Atlanta Biologics), 1% non-essential amino acids (Gibco), 1% sodium pyruvate (Gibco) and 1% Pen/Strep (Gibco). The HA-KPNA7 HEK293T stable cell line was generated by lentiviral transduction. Virus was produced by Fugene6 (Promega) mediated transfection of HEK293T cells with pWPI-HA-KPNA7, and the packaging plasmids psPAX2 (Addgene #12260) and pMD2.G (Addgene #12259). Viral supernatant was filtered (0.45 µm) prior to transduction. GFP-positive clonal lines were generated using the CellRaft AIR System (Cell Microsystems). U2OS cells (ATCC) cultured in DMEM high glucose (Gibco) with 10% FBS (Atlanta Biologics) and Pen/Strep (Gibco). Transfection of U2OS cells was performed with Fugene6 (Promega) per manufacturer's instructions. PC3(AR) cells were previously generated and stably express the androgen receptor (Jividen et al.,

2018). PC3(AR) cells were cultured in RPMI (Gibco) with 5% FBS (Atlanta Biologics) and Pen/Strep (Gibco).

H. *Heterokaryon cell fusion assay*

Heterokaryon assays were performed using methods previously described (Black et al., 2001). U2OS cells were transfected with HA-KPNA7 or HA-KPNA2 plasmids using Fugene6. After 24 hours, U2OS cells ($3x10^5$) and PC3(AR) ($3x10^5$) were co-seeded onto glass coverslips in 6 well dishes. After 16 hours 10 nM R1881 (synthetic androgen) was added to the culture to induce nuclear localization of the androgen receptor in PC3(AR) cells. U2OS-PC3(AR) fusion was induced by polyethylene glycol. Coverslips were washed with warm (37°C) PBS and placed cell side-down on warm (37°C) PEG-1500 (100 µl) for 30 seconds. PBS (37°C, 500 µl) was then added dropwise to the side of each coverslip to relieve surface tension. The coverslips were transferred to 6-well dishes, washed with warm (37°C) PBS, and finally placed in warm (37°C) media containing 10 µg/ml cycloheximide and 10 nM R1881. Plates were incubated at 37°C for 4 hours before processing for confocal immunofluorescence microscopy.

I. *Permeabilized cell import and export assays*

Nuclear import assays in digitonin-permeabilized cells were performed essentially as described (Cassany and Gerace, 2009). HEK293T cells seeded onto coverslips coated with 0.1 mg/ml poly-D-lysine at a density of $1.5x10^5$ cells/ml 24 h before use. Cells were washed 3x with ice-cold transport buffer (20 mM HEPES [pH 7.4], 110 mM potassium acetate, 2 mM magnesium acetate, 0.5 mM EGTA) containing 1 mM dithiothreitol (DTT) and 1 µg/ml portions (each) of leupeptin, pepstatin, and aprotinin and then permeabilized with 0.005% digitonin for 5 min on ice. Import reactions contained an energy-regenerating system (5 mg/ml of bovine serum albumin, 80 U of creatine phosphokinase/ml, 1.6 mg of creatine phosphate/ml, 1 mM ATP, 1 mM GTP), 1 µM NTF2 and 2 µM Ran, as well as 1 µM import substrate (GST-GFP fusion), 1 µM Importin-β and/or 1 µM MBP-KPNA7 as indicated. Reactions were carried out for 30 minutes at 30°C and terminated by transferring to ice-cold transport buffer. After 3 additional washes in transport buffer, coverslips were processed for confocal

immunofluorescence microscopy. Quantification of mean nuclear intensity was quantified using ImageJ software by making a mask with DAPI channel.

For export assays, a HEK293T cell line stably expressing HA-KPNA7 was generated by lentiviral transduction. Cells were plated and processed identically to the import assay for permeabilization and washing. Export mixtures were prepared with an energy-regenerating system (5 mg/ml of bovine serum albumin, 80 U of creatine phosphokinase/ml, 1.6 mg of creatine phosphate/ml, 1 mM ATP, 1 mM GTP), 1 μM NTF2 and 2 μM Ran, and CAS ranging in concentration from 0-2 μM as indicated. Export reactions were also incubated at 30°C for 30 min and for confocal immunofluorescence microscopy. Quantification of nuclear signal was performed using ImageJ software by making a mask with the DAPI channel.

J. *Confocal immunofluorescence microscopy*

Coverslips from import, export, heterokaryon, and transfection assays were fixed (3.75% formaldehyde, 15 min) and permeabilized (0.2% Triton x-100, 5 min). Coverslips were incubated in primary antibody diluted in IF microscopy blocking buffer (1X PBS, 2% BSA, 2% FBS) for ≥1 hour. After 3 PBS washes, the secondary antibodies were diluted in blocking buffer and incubated for 1 hour. DAPI was used to stain nuclei. Images were acquired by laser scanning confocal microscopy (Zeiss 880 LSM, Carl Zeiss) at 40×, 1.3 NA oil immersion objective with Zen software (Carl Zeiss). Quantification of nuclear signal was performed using ImageJ software.

K. *Antibodies*

Primary:

Anti-GST (B-14) – mouse monoclonal – Santa Cruz Biotechnology; sc-138

Anti-CAS (C-20) – mouse monoclonal – Santa Cruz Biotechnology; sc-271537

Anti-HA (16B12) – mouse monoclonal – Covance; MMS-101P

Anti-KPNA2 – rabbit polyclonal – abcam; ab84440

Anti-AR (AR21) – rabbit polyclonal – lab prepared to residues 1-21 of human AR (Jividen et al., 2018)

Anti-MBP – mouse monoclonal – NEB; E8032S

Anti-Imp-β (3E9) – mouse monoclonal – abcam; ab2811

Anti-KPNA7 – rabbit polyclonal – Millipore Sigma; HPA031395

Secondary:

Goat anti-mouse-IRDye-800 - Rockland; 610–132-121 (IB)

Donkey anti-rabbit-Alexa Fluor-680 – Invitrogen; A10043 (IB)

Goat Anti-Rabbit-Cy3 – Jackson ImmunoResearch; 115-165-144 (IF)

Donkey anti-mouse-FITC - Jackson ImmunoResearch; 715-225-151 (IF)

Goat anti-Mouse-Alexa Fluor-680 – Invitrogen; A21058 (IF)

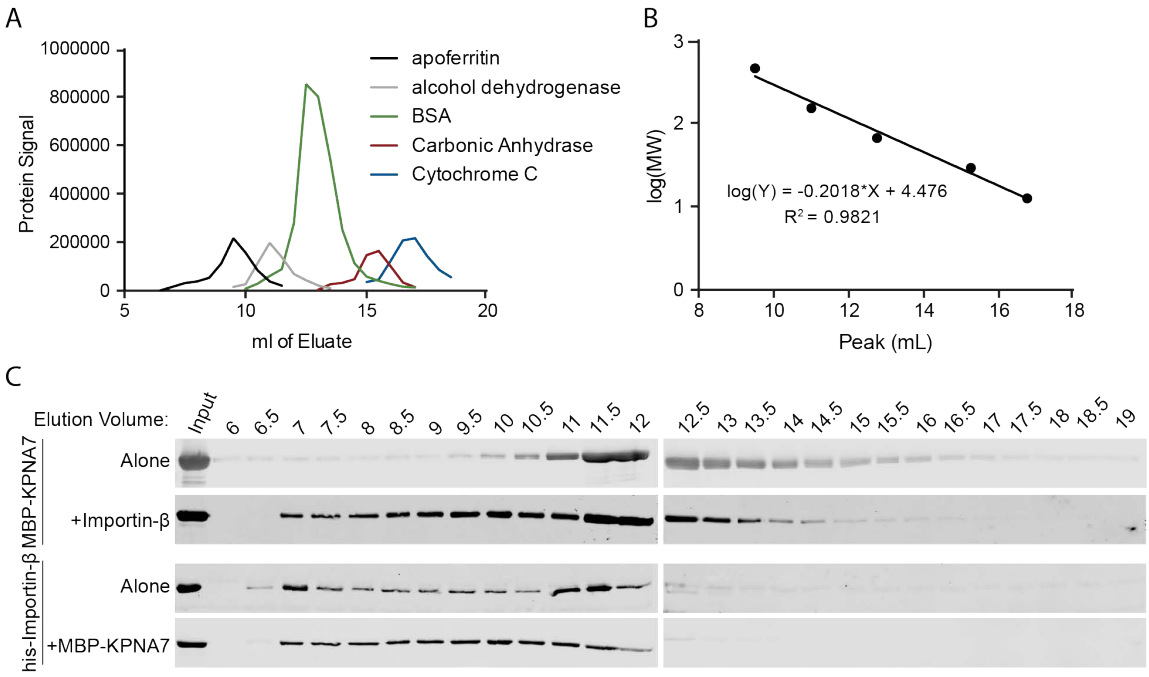


Figure S2.1: Superdex 200 size exclusion chromatography for characterization of the KPNA7-Imp-β complex.

(**A, B**) Elution profiles of protein standards and standard curve. (**C**) Immunoblotting detection of MBP-KPNA7 and his-Imp-β elutions from the Superdex200 size exclusion column. The values were used to the elution profiles (**Fig 1C**).

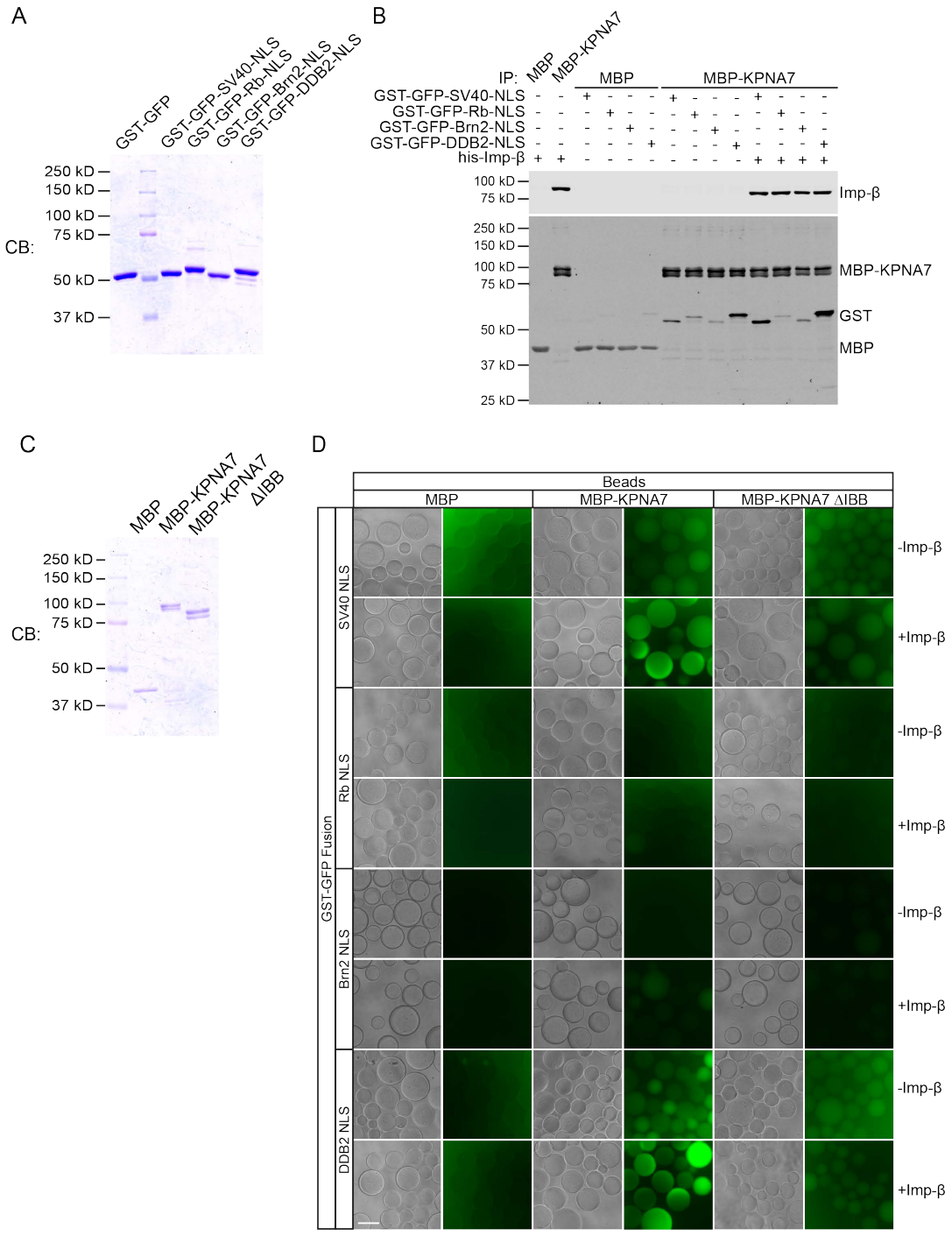


Figure S2.2: NLS Binding Assays with KPNA7.
(**A**) SDS-PAGE and Coomassie blue (CB) staining of GST-GFP-NLS fusions (1 µg each) used in this study. (**B**) Complete set of pull-down assays using MBP and MBP-KPNA7 immobilized on maltose beads (for **Fig. 2.8C**). (**C**) Maltose beads (0.5 µl loaded per lane) used for bead-based fluorescence binding assay showing the levels of immobilized protein (for **Fig. 2.9, 2.10**). (**D**) Complete set of bead-based fluorescence binding assay images, including phase contrast (for **Fig 2.9A**).

A

Species	Protein	Evidence	Experimental Method	Reference
Human	KPNA4	Specific NLS sequences (viral PB2) bind to full-length and ΔIBB KPNA4 with equal affinity; Mutational analysis of IBB domain recovers partial auto-inhibition	Pulldown (recombinant proteins)	Pumroy et al., 2015
Plasmodium faciparum	Pf-Impα	Pf-Impα and Pf-Impα-ΔIBB bind with equal affinity to TGS1-NLS; mutational analysis SKR motif recovers auto-inhibiton	Pulldown and Surface Plasmon Resonance (recombinant proteins)	Babar et al., 2016; Dey et al., 2018
Toxoplasma gondii	Tg-Impα	Full-length Tg-Impα binds to GCN5 in the absence of Impβ	Pulldown (in vitro translated proteins)	Bhatti and Sullivan, 2004
Arabidopsis thaliana	At-Impα	Impα binds to and transports NLS cargo in the absence of Impβ	ELISA-based binding assay; In vitro transport assay (recombinant proteins)	Hübner et al., 1999

B

```
KPNA4     MADNEKLDNQRLKNFKNKGRDLETMRRQRNEVVVELRKNKRDEHLLKRRNVPHEDICEDS   60
Tg-Impα   ------MERKLADRRSNFKKNFEDRRRKREDLQLQIRKTHREQNLAKKRAEALDAQDGR-   53
Pf-Impα   ------MDRRIEARRKEFKKNCDDTRRKREDLVVQIRKQQRECQLESKRAMVMANIGFE-   53
At-Impα   MSLRPNAKTEVRRNRYKVAVDAEEGRRRREDNMVEIRKSKREESLQKKRREGLQANQLP-   59
```

Figure S2.3: Importin-α paralogs with reduced auto-inhibition.
(**A**) Overview of studies suggesting modulation of auto-inhibition occurs with importin-α paralogs in diverse species. (**B**) Multiple sequence alignment showing sequence variations in the IBB domains from human (KPNA4), Toxoplasma gondii (Tg), Plasmodium falciparum (Pf), and Arabidopsis thaliana (At). Residues predicted to mediate auto-inhibition by binding to the minor and major NLS binding grooves are boxed.

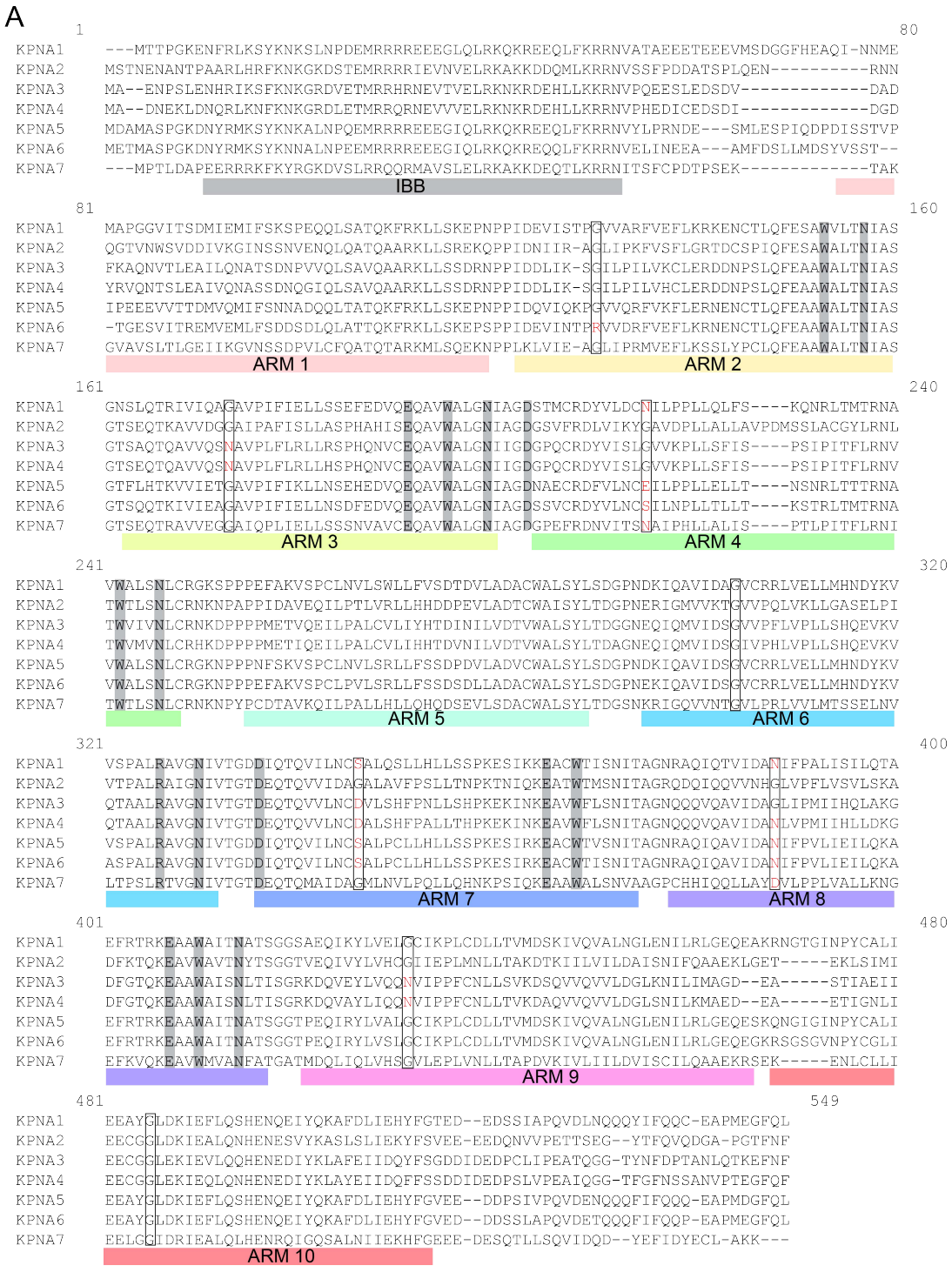


Figure S2.4: Multiple sequence alignment of KPNA proteins.
Alignment of the seven KPNA isoforms expressed in humans. Helix-breaking residues between helixes 1 and 2 of each ARM repeat are boxed. The helix-breaking residue is always glycine in KPNA2 (Pumroy et al., 2015), and typically a glycine in other KPNA members. Substitutions for glycine at the position are indicated (red). Conserved residues of the NLS binding groove are highlighted with a grey background.

Chapter III

An epilepsy-associated mutation in the nuclear import receptor KPNA7 reduces binding to nuclear localization signal sequences

I. Abstract

KPNA7 is a member of the Importin α family of nuclear import receptors which, in conjunction with Importin β, facilitate the translocation of signal containing proteins from the cytoplasm to the nucleus. The nuclear transport cycle is highly regulated, and disruption of transport often leads to disease states. Previously, mutations in KPNA7 have been associated with serve neurodevelopmental defect. Here, we have performed biochemical characterization of two disease-associated mutations in KPNA7 and determined the amino acid substitution E344Q reduces the capacity of KPNA7 for binding and facilitating nuclear import of both monopartite and bipartite nuclear localization signals. We reasoned that reduced KPNA7 activity toward neuron-specific protein cargoes could be the basis for the mutant contribution to the disease state and identified KPNA7 interacting proteins in induced pluripotent stem cell-derived neurons by protein affinity chromatography and SILAC-based mass spectrometry. The heterogeneous nuclear ribonucleoproteins, hnRNP R and hnRNP U were identified as KPNA7 interacting proteins, and binding and transport of the nuclear localization signal sequences from each protein were reduced by the E344Q substitution in KPNA7. Finally, a secondary effect of the mutation which causes the E344Q substitution was identified whereby disruption of binding of the CCCTC-binding protein, CTCF, to exon 7 of the *KPNA7* gene may impact regulation of gene expression.

II. Introduction

Maintenance of cellular hemostasis is critically dependent on the bi-directional transport of proteins and RNAs between the cytoplasm and nucleus. Dysregulation of protein transport has been implicated in several diseases including cancer and diverse neurodegenerative disorders (Kau et al., 2004; Patel and Chu, 2011). This includes disruption of pathways responsible for facilitating nuclear import of proteins containing nuclear localization signal (NLS) sequences. The composition and order of NLS sequences vary but primarily consist of either one (monopartite) or two (bipartite) clusters enriched in the basic amino acids lysine and arginine (Dingwall and Robbins, 1988; Kalderon et al., 1984). Classical NLS import is mediated by Importin α (Imp-α) and Importin β (Imp-β) (Pemberton and Paschal, 2005). Imp-α functions as an adapter and makes direct contact with both an NLS sequence and Imp-β to facilitate the assembly of a heterotrimeric import complex that shuttles from the cytoplasm to the nucleus (Damelin et al., 2002). In the nucleoplasm, a high nuclear concentration of RanGTP facilitates the release of the NLS containing protein and recycling of the import machinery (Görlich et al., 1995; Weis, 2003).

The human Imp-α family, also known as the karyopherin α family, contains seven members named karyopherin α1-α7 (KPNA1-KPNA7) (Kelley et al., 2010). These receptors share a highly similar structure including an N-terminal Imp-β binding (IBB) domain, and two NLS binding surfaces formed by ten armadillo (ARM) repeats (Conti et al., 1998; Kobe, 1999). The major NLS binding groove (ARMs 2-4) is utilized for binding both mono- and bipartite NLS sequences, while the minor groove (ARMs 6-8) is primarily used for binding the smaller, second cluster of basic residues in bipartite NLS sequences (Fontes et al., 2003) Despite similar structures, and high conservation of the amino acids residues that make up the NLS-binding groove, there is discrimination in the specific NLS sequences each isoform can bind (Köhler et al., 1999; Pumroy and Cingolani, 2015).

NLS cargo specificity, combined with differential expression of the Imp-α isoforms, termed Importin-α isoform switching, is utilized in developmental programs across species (Pumroy and Cingolani, 2015). Examples include germ cell maturation in

Drosophila and mice, (Hogarth et al., 2007; Mason et al., 2002; Mihalas et al., 2015), and neuronal development in mice (Yasuhara et al., 2006). Changes in the temporal expression or function of individual Imp-α isoforms have been implicated in multiple diseases including cancers (Jensen et al., 2011; Kim et al., 2000; Wang et al., 2011), neurodegenerative disorder (Lee et al., 2006; Morris et al., 2012; Paciorkowski et al., 2014; Zhang et al., 2006) and inflammatory bowel disease (Theiss et al., 2009). These data support the view that differential expression of KPNA proteins can be used in a regulatory capacity. This presumably relates to the fact that at least some NLS cargo proteins are preferentially transported by select Imp-α isoforms (Yasuhara et al., 2013).

KPNA7 is the most recently identified, and also most divergent member of the Imp-α family in humans (Kelley et al., 2010). It shares 55% amino acid identity with its closest related isoform, KPNA2 (Kelley et al., 2010). In multiple species, expression of KPNA7 orthologs is mainly limited to the ovary, oocyte and developing embryo (Hu et al., 2010; Tejomurtula et al., 2009; Wang et al., 2014, 2012). In humans, high KPNA7 expression has been observed in pancreatic cancers as a result of gene duplication events (Vuorinen et al., 2018). Additionally, compound heterozygous mutations in KPNA7 have been associated with neurodevelopmental defect including severe developmental disability, infantile spasms, and epilepsy (Paciorkowski et al., 2014). The two mutations (c.1015C>G and c.1030G>C) both result in amino acid substitutions (P339A and E344Q) in the seventh ARM repeat of the protein, proximal to the minor NLS binding groove, but how these amino acid changes affect KPNA7 protein function has not been directly tested. Regardless, the association of KPNA7 mutation with neuronal defect suggests KPNA7 may make critical contributions to the development of the nervous system.

In this study, we have identified and induction of KPNA7 expression during neuronal differentiation of human induced pluripotent stem cells (iPSCs). Based on this discovery we utilized iPSC-derived neurons to discover neuronal KPNA7 binding partners and identified two heterogeneous nuclear ribonucleoproteins, hnRNP R and hnRNP U, as KPNA7 interacting proteins. KPNA7 binds to and facilitates nuclear import of a functional bipartite NLS sequence in hnRNP R and a function monopartite NLS sequence in hnRNP U. We observed a reduction in binding by the epilepsy-associated

KPNA7(E344Q) to both full-length hnRNP R and hnRNP U, and to the isolated NLS sequences from each protein. In addition, the E344Q substitution significantly reduces KPNA7-mediated nuclear import of the two NLS sequences in permeabilized-cell transport assays. Finally, the causal mutation of the E344Q substitution, c.1030G>C, occurs in a CTCF binding-motif in exon seven of the *KPNA7* gene and significantly reduces CTCF affinity for the *KPNA7* DNA sequence.

III. Results

A. *The amino acid substitution E344Q in KPNA7 reduces NLS binding and import*

The auto-inhibitory function of the IBB domain of Imp-α is a key regulatory feature. This N-terminal domain has an NLS-like sequence that functions to mask the NLS binding groove. The receptor opens via concerted binding of both an NLS containing cargo protein and Imp-β. We have previously shown that KPNA7 is not subject to auto-inhibition in the same manner as other Imp-α isoforms but instead adopts a predominately open-state, allowing the receptor to bind Imp-β in the absence of an NLS cargo (Chapter II). The locations of the epilepsy-associated substitutions (P339A and E344Q) in the NLS binding groove of KPNA7 (**Fig 3.1A**) suggest they may impact both the NLS binding function of the receptor as well as IBB binding and therefore the auto-inhibitory state. We assayed the binding of recombinant wild-type and mutant KPNA7 to Imp-β in the absence of an NLS cargo to determine if the amino acids substitutions affect receptor auto-inhibition (**Fig 3.1B**). Similar levels of Imp-β binding are observed for wild-type KPNA7 and the mutants which suggest the open-state of the receptor is retained. Additionally, we performed CAS-mediated nuclear export assays and found minimal differences in nuclear export between the 3 forms of KPNA7 (**Fig S3.1**). Together these data suggest the open-state and nuclear localization of KPNA7 are unaffected by the epilepsy-associated mutations.

The amino acid substitution E354Q decreases binding of the Imp-α isoform KPNA2 to the bipartite NLS sequence from retinoblastoma protein (Rb) by approximately half (Paciorkowski et al., 2014). This substitution of the conserved E354 residue is homologous to the E344Q substitution in KPNA7, and this result suggests the

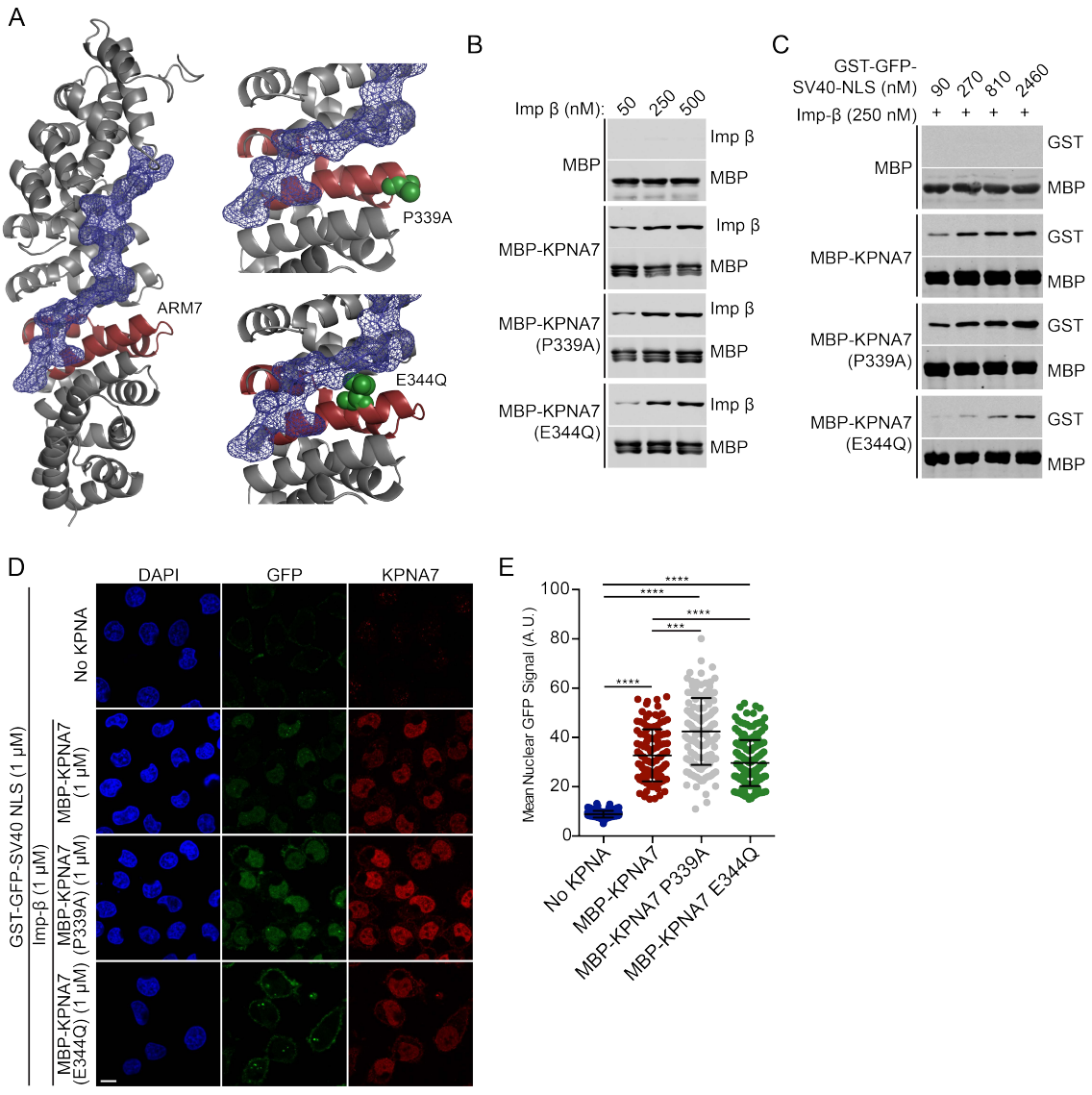


Figure 3.1: The amino acid substitution E344Q in ARM7 of KPNA7 reduces binding to the SV40 NLS and reduces import in permeabilized cells.
(A) Structural model of KPNA7 generated by threading the KPNA7 primary sequence onto the crystal structure of KPNA2 (PBD: 1PJM) using the SWISS-MODEL software. The bipartite NLS sequence from Rb is shown in blue wire mesh. Highlighted in red is the 7th ARM repeat. Insets show the P339A and E344Q amino acid substitutions in green space-fill. Images rendered in PyMol. **(B)** MBP fusion proteins were immobilized on amylose resin and incubated with the indicated concentrations of Imp-β. **(C)** MBP fusion proteins were immobilized on amylose resin and incubated with GST-GFP-SV40-NLS at the indicated concentrations and Imp-β (250 nM). **(D)** Fluorescence microscopy images of permeabilized cell import assays showing the nuclear accumulation of GST-GFP-SV40 NLS. The assays were performed in HeLa cells permeabilized with 0.005% digitonin, and incubated with Ran (0.2 μM), NTF2 (2 μM) and an energy regenerating system with the recombinant factors indicated. Confocal IF microscopy was used to visualize. **(E)** Quantification of SV40-NLS import for each reaction. At least 100 cells of each were quantified. (Scale: 10 μm. **** $p<0.0001$, *** $p<0.001$)

substitution in KPNA7 would have a similar effect on NLS binding. The effects of both the E344Q and P339A substitutions in KPNA7 on NLS binding and transport have not been previously examined. We evaluated the effect of the substitutions on binding the prototypical monopartite NLS from the SV40 Large-T antigen (PKKKRKV) in a binding assay using a recombinant GST-GFP-SV40 NLS fusion (**Fig 3.1C**). In the presences of Imp-β (250 nM), KPNA7(E344Q) binding to the SV40 NLS is reduced compared to wild-type KPNA7. To determine if this reduction affects SV40 NLS import by KPNA7 we performed digitonin-permeabilized cell import assays with recombinant proteins (**Fig 3.1D**). No import of the SV40 NLS is seen in the absence of KPNA7. KPNA7 and Imp-β addition results in nuclear accumulation of the NLS construct for wild-type and both KPNA7 mutants (**Fig. 3.1D, E**). Compared to wild-type KPNA7, however, KPNA7(E344Q) facilitates significantly less SV40 NLS import (~20% reduction, $p<0.0001$). These data suggest that the E344Q substitution in the KPNA7 minor NLS binding groove reduces the receptors function toward the monopartite SV40 NLS sequence.

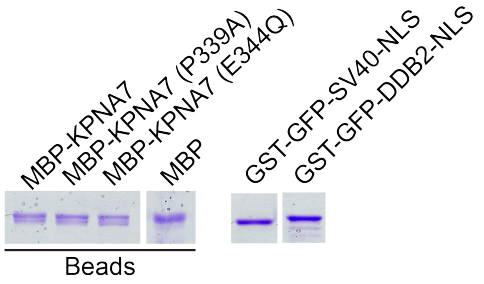
As a method of rapidly evaluating low-affinity NLS binding by KPNA7, we utilized a bead-based fluorescence assay in which steady-state association of a GST-GFP-NLS cargo to bead-bound KPNA7 is observed by fluorescence microscopy (Patel and Rexach, 2008). We previously adapted this assay to use fluorescently labeled MBP as an internal control for semi-quantitative measurements (Chapter II). Briefly, MBP-KPNA7 is bound to amylose beads and mixed with beads containing unfused MBP and a tracer amount of MBP labeled with IRDye-680. This bead mixture is incubated with a GST-GFP-NLS cargo protein and association of NLS to MBP-KPNA7 beads and control beads is evaluated by fluorescence microscopy. We utilized this assay to evaluate binding of the KPNA7 mutants to both the monopartite SV40 NLS sequence and the bipartite NLS sequence from the DNA-damage binding protein, DDB2, and the dependence of this binding on addition of Imp-β (**Fig. 3.2A-C**). Both MBP-KPNA7 and MBP-KPNA7(P339A) beads specifically bind both NLS sequences and Imp-β increases the magnitude of binding (**Fig 3.2D**). In contrast, minimal binding of either NLS sequence is observed with MBP-KPNA7(E344Q) beads even with the addition of Imp-β to the

A

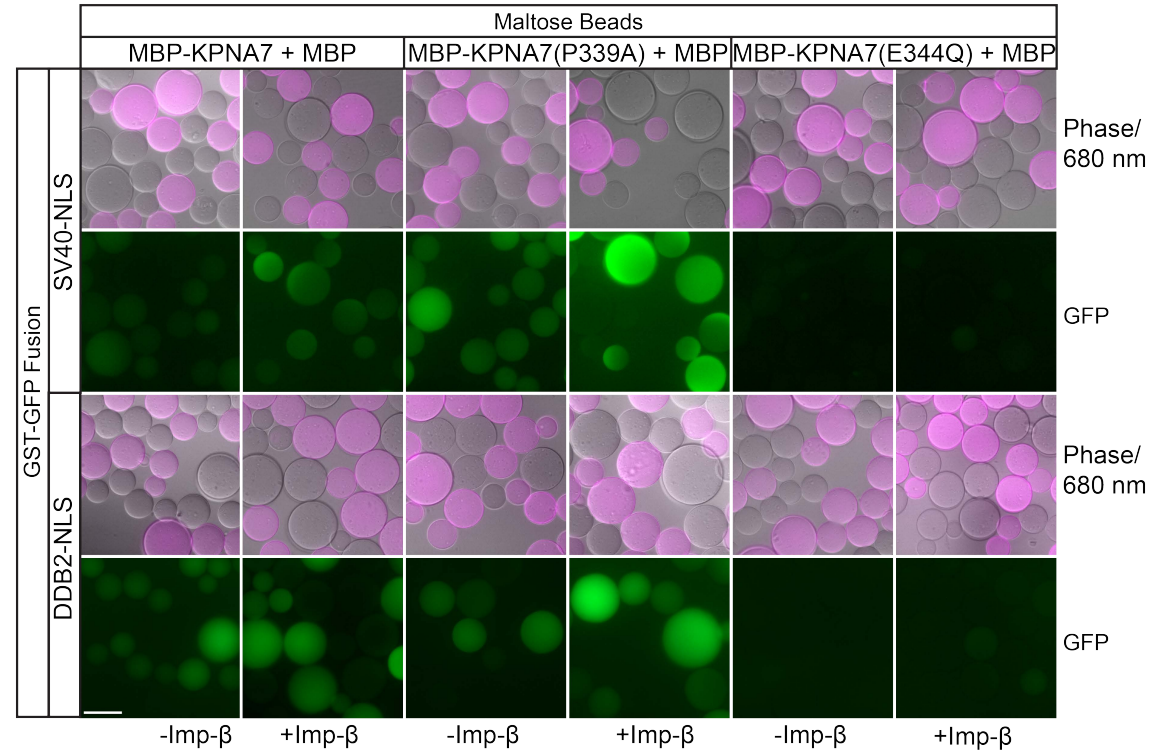
NLS Sequences

SV40	PKKKRKV
DDB2	RPRNKRSRSPLELEPEAKKLCAK

B



C



D

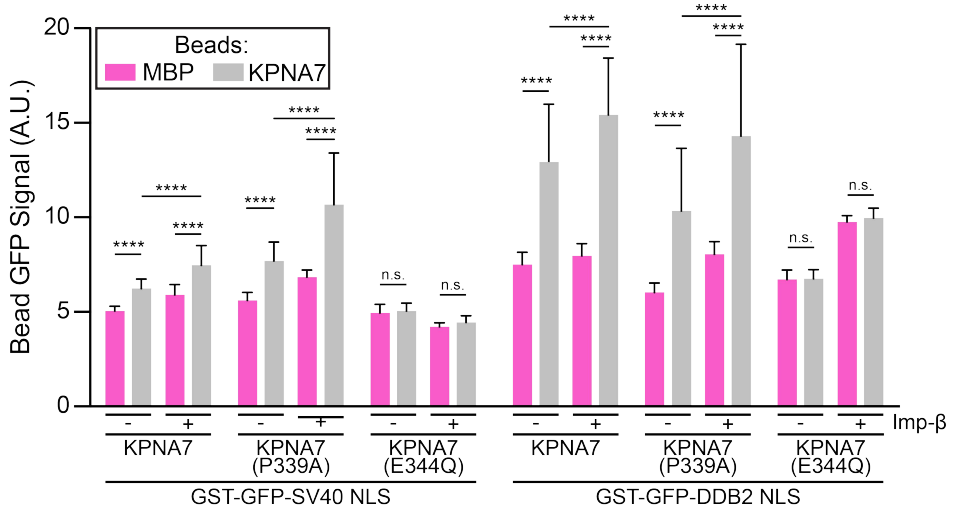
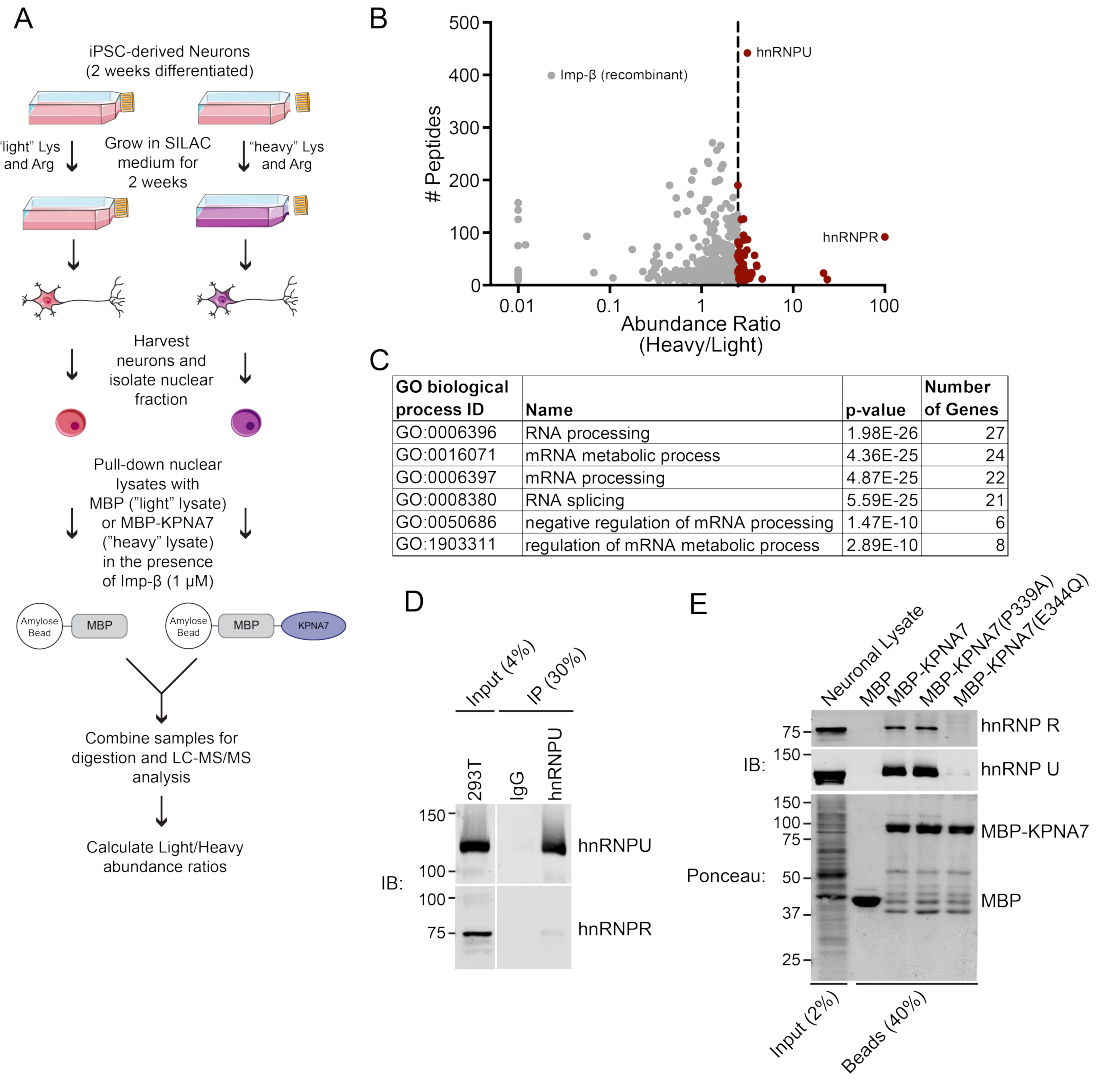


Figure 3.2: E344Q substitution reduces binding to mono- and bipartite NLS sequences by bead-based fluorescence imaging.
(A) NLS sequences from SV40 and DDB2. **(B)** Amylose beads (0.5 µl per lane) used for fluorescence binding assay analyzed by SDS-PAGE and Coomassie blue (CB) staining. **(C)** Bead-based fluorescence imaging assay of MBP-KPNA7 proteins to the indicated NLS sequence. MBP-KPNA7 beads were combined with control MBP beads with a tracer amount of IRDye680-labeled MBP. Incubations were performed with the GST-GFP NLS constructs (750 nM) ± Imp-β (1 µM), washed and imaged. **(D)** Fluorescence intensity of GFP was measured for MBP beads (680 nm positive) and MBP-KPNA7 beads (680 nm negative), and mean values were plotted. At least 30 beads of each condition were measured. (Scale: 100 µM, ****$p<0.0001$)

reaction (**Fig. 3.2D**). We conclude that the E344Q substitution in KPNA7 significantly reduces binding to both mono and bipartite NLS sequences.

B. *Identification of KPNA7 interacting proteins in neurons.*

NLS containing proteins which interact with KPNA7 in neurons have not previously been investigated. Reduced NLS binding by KPNA7(E344Q) suggests the identification of KPNA7 cargo proteins would help advance our understanding of KPNA7 function and generate hypotheses for how KPNA7 mutants affect neuronal differentiation. To identify KPNA7 interacting proteins in human induced pluripotent stem cell (iPSC)-derived neurons we utilized stable isotope labeling by amino acids in cell culture (SILAC) and mass spectrometry (**Fig. 3.3A**). Briefly, iPSC-derived neurons were differentiated in neuron media for 2 weeks and then transferred to SILAC neuron media containing light, or heavy, isotopes of the amino acids lysine and arginine (13C615N2-lysine/13C615N4-arginine (Lys8/Arg10)) for an additional two weeks. The four-week neurons were harvested, nuclear fractions isolated, and the light and heavy lysates were pulled-down with MBP or MBP-KPNA7 beads respectively (**Fig. 3.3A**). The samples were then combined and analyzed by liquid chromatography with tandem mass spectrometry (LC-MS/MS). In total 1257 proteins were identified and abundance ratios (Heavy/Light) were calculated to determine if a given protein was enriched in the MBP-KPNA7 pulldown. There were 39 proteins with abundance ratios greater than 2.75 and more than 10 identified peptides (**Table S3.1**) which are shown in red in the abundance ratio plot (**Fig. 3.3B**). We analyzed these proteins for enrichment of GO biological processes and found a significant enrichment in processes relating to RNA regulation and splicing (**Fig. 3.3C**). Additionally, we utilized cNLS Mapper to search for NLS sequences in these proteins and found 29 out of 39 contained putative NLS sequences. Previous studies investigating KPNA7 interacting proteins in HEK293 and pancreatic cancer cell lines have also identified highly significant enrichment for proteins involved in RNA processing (Kimoto et al., 2015; Vuorinen et al., 2017). In particular, multiple hnRNP proteins were identified as candidate binding proteins for KPNA7 in pancreatic cancer cells (Vuorinen et al., 2017). Peptides from validated binding partners from these studies such as MVP, ZNF414, and DDB2 were not identified in our screen which may reflect either low expression levels or lack of nuclear enrichment in iPSC-derived neurons.



GO biological process ID	Name	p-value	Number of Genes
GO:0006396	RNA processing	1.98E-26	27
GO:0016071	mRNA metabolic process	4.36E-25	24
GO:0006397	mRNA processing	4.87E-25	22
GO:0008380	RNA splicing	5.59E-25	21
GO:0050686	negative regulation of mRNA processing	1.47E-10	6
GO:1903311	regulation of mRNA metabolic process	2.89E-10	8

Figure 3.3: SILAC-Mass Spectrometry identifies hnRNPR and hnRNPU as candidate KPNA7 interacting proteins.
(A) Schematic for the SILAC-mass spectrometry experiment to identify nuclear KPNA7 interacting proteins from iPSC-derived neurons. **(B)** Plot showing the abundance ratio (measure of the enrichment for the heavy KPNA7 pull-down sample) versus the number of peptides identified from a given protein. A line is drawn at abundance ratio of 2.75 and proteins with greater than this value are shown in red. **(C)** Gene ontology (GO) biological process enrichment analysis was performed on the 41 genes with greater than 10 identified peptides and abundance ratio greater than 2.75. **(D)** Immunoprecipitation of hnRNP U from HEK293T cells. Co-IP of hnRNP R was analyzed by western blot. **(E)** Validation of two candidate binding partners, hnRNP R and hnRNP U. MBP and MBP-KPNA7 proteins were immobilized on amylose resin and used to pull-down 4-week old 9429A neuronal lysate. Binding of hnRNP R and hnRNP U was analyzed by western blot.

C. *Characterization of hnRNP R and hnRNP U*

Two heterogeneous nuclear ribonuclear proteins, hnRNP R and hnRNP U, had both large numbers of identified peptides (92 and 442 respectively) and a high abundance ratio (100 and 3.159 respectively) which is indicative of enriched binding to KPNA7. The hnRNP family of proteins is a group of at least 20, primarily nuclear, RNA binding proteins which form ribonucleoprotein complexes in cells. (Dreyfuss et al., 1993; Han et al., 2010). While other members of the hnRNP family were identified in our screen none had abundance ratios of greater than 2.75. Both hnRNP R and hnRNP U were initially identified as part of a large supramolecular heterogeneous ribonucleoprotein particle, but have also been shown to be part of other smaller ribonucleoprotein complexes with functions in both the nucleus and cytoplasm (Chun et al., 2016; Fukuda et al., 2009; Lee et al., 2015; Piñol-Roma et al., 1988). One possible explanation for the identification of multiple hnRNPs is an interaction of KPNA7 with a single hnRNP particle containing multiple proteins. To determine if there may be a direct interaction of hnRNP and hnRNP U, we performed immunoprecipitation of hnRNP U in HEK293T cells (**Fig 3.3D**). Minimal hnRNP R co-immunoprecipitated with hnRNP U suggesting that the two proteins are not in the same hnRNP complex. We further substantiated this result by performing size exclusion chromatography (SEC) of HEK293T cell lysate with a Superose6 column (**Fig S3.2**). hnRNP U, which is around 120 kD by gel electrophoresis, elutes mainly as part of a complex between 200 and 450 kD, however, a slightly larger isoform appears to be a part of larger complexes. In contrast, hnRNP R, which is about 80 kD by gel electrophoresis, has an apparent size of greater than 1 MD by SEC.

D. *KPNA7 binding to hnRNP R and hnRNP U is reduced by the E344Q substitution*

We validated the specific interaction of KPNA7 with both hnRNP R and hnRNP U by utilizing recombinant MBP and MBP-KPNA7 to pull-down iPSC-derived neuronal lysate and analyzed the binding of the proteins of interest by blotting (**Fig 3.3E**). Wild-type KPNA7 bound strongly to both hnRNP R and hnRNP U, while no background binding to MBP was observed. In this experiment, we also interrogated the binding of both epilepsy-associated KPNA7 mutants. We observed a striking reduction in pull-down of both hnRNP R and hnRNP U by the E344Q mutant. Binding to the P339A mutant, however,

was similar to wild-type KPNA7. Our previous data demonstrated the reduced function of KPNA7(E344Q) for NLS binding which suggests the observed interaction of KPNA7 and the hnRNP proteins is via NLS binding.

E. *hnRNP R contains a functional bipartite NLS sequence*

hnRNP R is made up of multiple domains which include three RNA-recognition motifs and an RNA binding RGG (Arg-Gly-Gly) box (**Fig 3.4A**). Two putative NLS sequences in hnRNP R have been reported based on sequence analysis and similarity to a closely related hnRNP, hnRNP Q (Cappelli et al., 2018; Mourelatos et al., 2001). These include a monopartite NLS (mNLS, a.a. 412-418) and a bipartite NLS (bNLS, a.a. 565-589) (**Fig 3.4A**). Interrogation of the hnRNP R amino acid sequence with the cNLS mapper algorithm (Kosugi et al., 2009b) also identifies each of these regions as potential NLS sequences. While no direct interrogation of the sufficiency and necessity of each of the NLS sequences has been performed, analysis of the localization of a truncated isoform of hnRNP Q, which lacks the homologous bipartite NLS sequence and is primarily cytoplasmic in localization, suggests this NLS is responsible for the nuclear localization of the receptor (Cappelli et al., 2018). To evaluate the sufficiency of each NLS sequence in promoting nuclear localization of hnRNP R, we generated mutant constructs by mutating key basic residues to alanine (mutated residues are shown in bold in **Fig 3.4A, B**). Each mutant construct was expressed in HeLa cells and cellular localization was examined by immunofluorescence confocal microscopy (**Fig 3.4C**). We quantified the ratio of nuclear-to-cytoplasmic (N:C) localization of each construct and found the wild-type and monopartite NLS mutant, $AAA^{414\text{-}416}$, to have a higher nuclear signal (an N:C ratio of greater than 1) in 100% of cells (**Fig 3.4D**). In contrast, both bipartite mutants, $AAA^{572\text{-}574}$ and $AAA^{584\text{-}586}$, were distributed between the nucleus and cytoplasm and more than 90% had an N:C ratio of less than 1. This result indicates that the bipartite NLS in hnRNP R alone is sufficient to drive complete nuclear localization, while the monopartite NLS is not. To further evaluate the functionality of each proposed NLS sequence we expressed each as a GST-GFP fusion protein. We utilized these proteins in a pull-down binding assay with the seven Imp-α isoforms expressed as $^{35}S^{-}$methionine labeled proteins by in vitro transcription and translation (**Fig 3.4E**). We observed no binding to the proposed monopartite NLS sequence by any Imp-α isoform. In contrast,

multiple Imp-α isoforms, including KPNA7, bound to the proposed bipartite NLS sequence. From these data, we conclude that hnRNP R contains a functional bipartite NLS sequence (a.a. 565-589) which directs its localization to the nucleus.

F. *KPNA7 binds to the hnRNP R bipartite NLS*

We next evaluated the effect of the hnRNP R NLS mutations on the interaction with KPNA7. We expressed the mutant hnRNP R constructs in HEK293T cells which stably express HA-KPNA7 and performed an HA immunoprecipitation (**Fig 3.4F**). We observed co-immunoprecipitation of wild type hnRNP R with KPNA7. After normalization to protein input levels, the two bipartite NLS mutants, $AAA^{572\text{-}574}$ and $AAA^{584\text{-}586}$, reduced binding to approximately 20% of the wild type protein, similar to background binding observed in controls cells without HA-KPNA7. We also observed a greater than two fold-reduction by the $AAA^{414\text{-}416}$ monopartite NLS mutant. While we did not see an effect of mutating this putative NLS on protein localization, nor binding to the isolated NLS sequence by Imp-α isoforms, we cannot rule out a contribution of this sequence on KPNA7 binding to full-length hnRNP R. Nonetheless, these results suggest that the interaction between KPNA7 and hnRNP R is primarily through the functional bipartite NLS sequence.

G. *KPNA7 binding to the bipartite NLS sequence in hnRNP R is reduced by the E344Q substitution*

We next characterized the impact of the epilepsy-associated mutations in KPNA7 specifically on the interaction with the hnRNP R NLS sequences. We utilized recombinant MBP-KPNA7 fusion proteins to perform a binding assay with GST-GFP-hnRNP R NLS constructs (**Fig 3.5A**). Strong binding of MBP-KPNA7 and MBP-KPNA7(P339A) to the hnRNP R bipartite NLS is observed, while MBP-KPNA7(E344Q) shows reduced binding (more than 10-fold less than wild-type KPNA7). Minimal binding to the hnRNP R mNLS sequence is observed for any of the forms of KPNA7.

To determine if KPNA7 can facilitate nuclear import of the two hnRNP R NLS sequences, we performed digitonin-permeabilized cell import assays (**Fig 3.5B**). We observed increased nuclear GFP signal for both the monopartite and bipartite NLS with the addition of each KPNA7 construct relative to the no KPNA control (**Fig 3.5C**). There

is no significant difference in the nuclear accumulation of hnRNP R-mNLS between each of the KPNA7 constructs. KPNA7(E344Q) facilitates significantly less import hnRNP R-

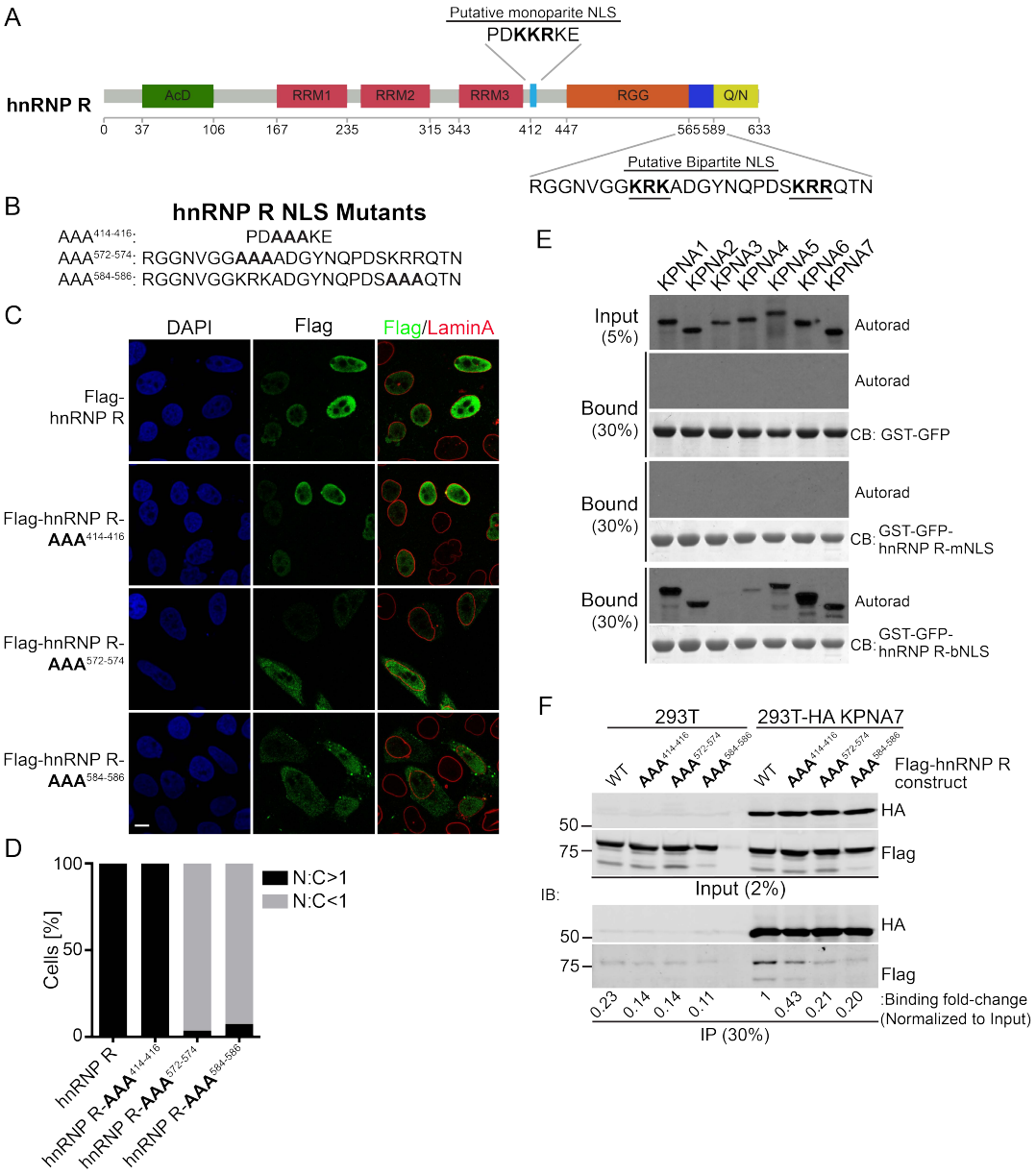


Figure 3.4: hnRNPR contains a functional bipartite NLS.
(A) Representation of the hnRNP R protein. The domain labeled are as follows: AcD – Acidic Domain, RRM – RNA recognition motif, RGG – RNA-binding Arg-Arg-Gly box, Q/N – Glu and Asn rich domain. Putative mono and bipartite NLS are indicated. **(B)** Mutant constructs were made with the indicated residues substituted with Alanine. **(C)** The NLS mutant constructs were expressed in HeLa cells and image by confocal immunofluorescence microscopy. Nuclear and cytoplasmic Flag intensity was quantified and the percentage of cells with nuclear-to-cytoplasmic (N:C) ratios of greater than or less than 1 is plotted in **(D)**. **(E)** Putative monopartite and bipartite NLS were expressed as GST-GFP fusion proteins and immobilized on glutathione beads for use in a binding assay with ^{35}S-labeled KPNA proteins. Bound fractions were analyzed by SDS-PAGE and autoradiography. **(F)** An HA-immunoprecipitation was performed in HEK293T and HEK293T(HA-KPNA7) cells transfected with the indicated hnRNP R constructs and analyzed by SDS-PAGE and western blot. Binding was quantified by Flag antibody and normalized to input. Binding fold-change relative to WT hnRNP R binding to HA-KPNA7 is below each lane.

bNLS compared to wild-type KPNA7 (15% reduction, $p<0.0001$). These data indicate that KPNA7 binds to, and facilitates nuclear import of, the hnRNP R bipartite NLS and that the E344Q substitution reduces this function.

H. *KPNA7 binding to the monopartite NLS sequence in hnRNP U is reduced by the E344Q substitution*

hnRNP U is structurally distinct from hnRNP R. The protein is unique among the hnRNP family in that it does not contain an RNA recognition motif, but instead binds RNAs only through the arginine and glycine-rich RGG box at the C-terminus (Kiledjian and Dreyfuss, 1992) (**Fig 3.6A**). Interestingly, while a monopartite NLS sequence was identified in hnRNP U and the protein is strongly nuclear in localization, early heterokaryon studies determined that hnRNP U is confined to the nucleus and does not undergo nucleocytoplasmic shuttling in HeLa cells (Eggert et al., 1997; Kiledjian and Dreyfuss, 1992; Piñol-Roma and Dreyfuss, 1992). Regardless, the nuclear localization of the protein was shown to be dependent on the NLS sequence (Eggert et al., 1997). We expressed the monopartite NLS sequence from hnRNP U as a GST-GFP fusion for use in binding and transport assays. In the presence of Imp-β, all three KPNA7 proteins bound the hnRNP U-mNLS sequence (**Fig 3.6B**). Similar levels of binding were observed between wild-type KPNA7 and the P339A mutant. There was a large (~10-fold) reduction in binding to the hnRNP U-mNLS by the E344Q mutant.

To determine if KPNA7 can facilitate import of the hnRNP U NLS sequence, we again performed digitonin-permeabilized cell import assays (**Fig 3.6C**). Significant, KPNA7-dependent nuclear accumulation of the GST-GFP-hnRNP U-mNLS sequence was observed (**Fig 3.6D**). Consistent with our observations of other NLS sequences, KPNA7(E344Q)-dependent import was significantly reduced (~30% reduction, $p<0.0001$) compared to wild-type KPNA7. These data further demonstrate that the E344Q substitution in KPNA7 reduces the transport function of the receptor for both monopartite and bipartite NLS sequences, including those which are bound specifically in neurons.

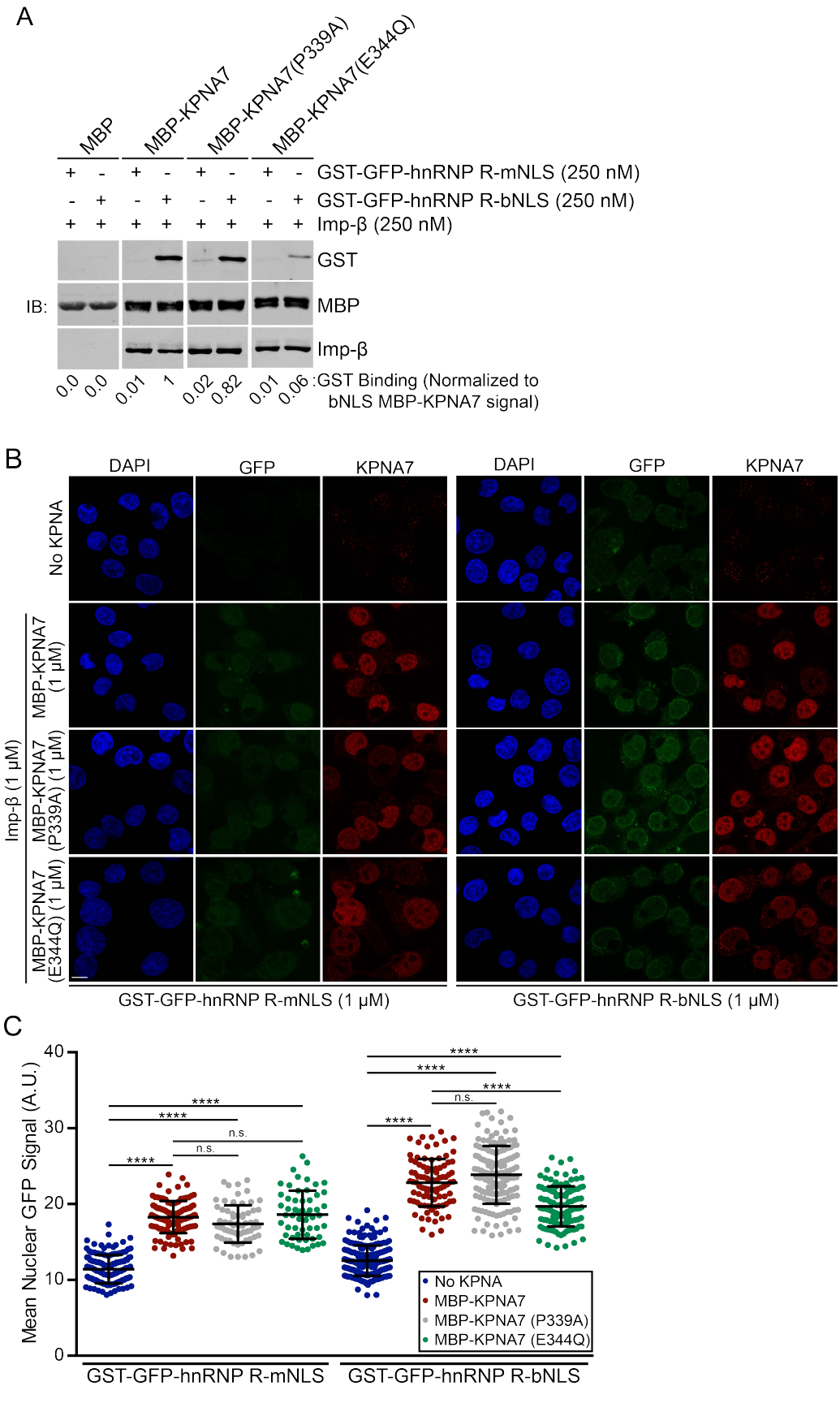

A
MBP
MBP-KPNA7
MBP-KPNA7(P339A)
MBP-KPNA7(E344Q)
GST-GFP-hnRNP R-mNLS (250 nM)
GST-GFP-hnRNP R-bNLS (250 nM)
Imp-β (250 nM)
IB:
GST
MBP
Imp-β
0.0 0.0 0.01 1 0.02 0.82 0.01 0.06 :GST Binding (Normalized to bNLS MBP-KPNA7 signal)
B
DAPI
GFP
KPNA7
No KPNA
MBP-KPNA7 (1 μM)
MBP-KPNA7 (P339A) (1 μM)
MBP-KPNA7 (E344Q) (1 μM)
Imp-β (1 μM)
GST-GFP-hnRNP R-mNLS (1 μM)
GST-GFP-hnRNP R-bNLS (1 μM)
C
Mean Nuclear GFP Signal (A.U.)
0
10
20
30
40
****
n.s.
No KPNA
MBP-KPNA7
MBP-KPNA7 (P339A)
MBP-KPNA7 (E344Q)
GST-GFP-hnRNP R-mNLS
GST-GFP-hnRNP R-bNLS

Figure 3.5: KPNA7 binding to the bipartite NLS of hnRNPR is reduced by the E344Q substitution.
(A) Binding assay with the indicated MBP proteins bound to amylose resin and the indicated GST-GFP hnRNP R NLS constructs (250 nM) in the presence of Imp-β (250 nM). NLS binding was normalized to MBP signal and plotted below each lane. **(B)** Fluorescence microscopy images of permeabilized cell import assays with GST-GFP-hnRNP R NLS constructs. The assays were performed in HeLa cells permeabilized with 0.005% digitonin, and incubated with Ran (0.2 μM), NTF2 (2 μM) and an energy regenerating system with the recombinant factors indicated. Confocal IF microscopy was used to visualize. **(C)** Quantification of NLS import for each reaction. At least 100 cells of each were quantified. (Scale: 10 μm. **** $p<0.0001$)

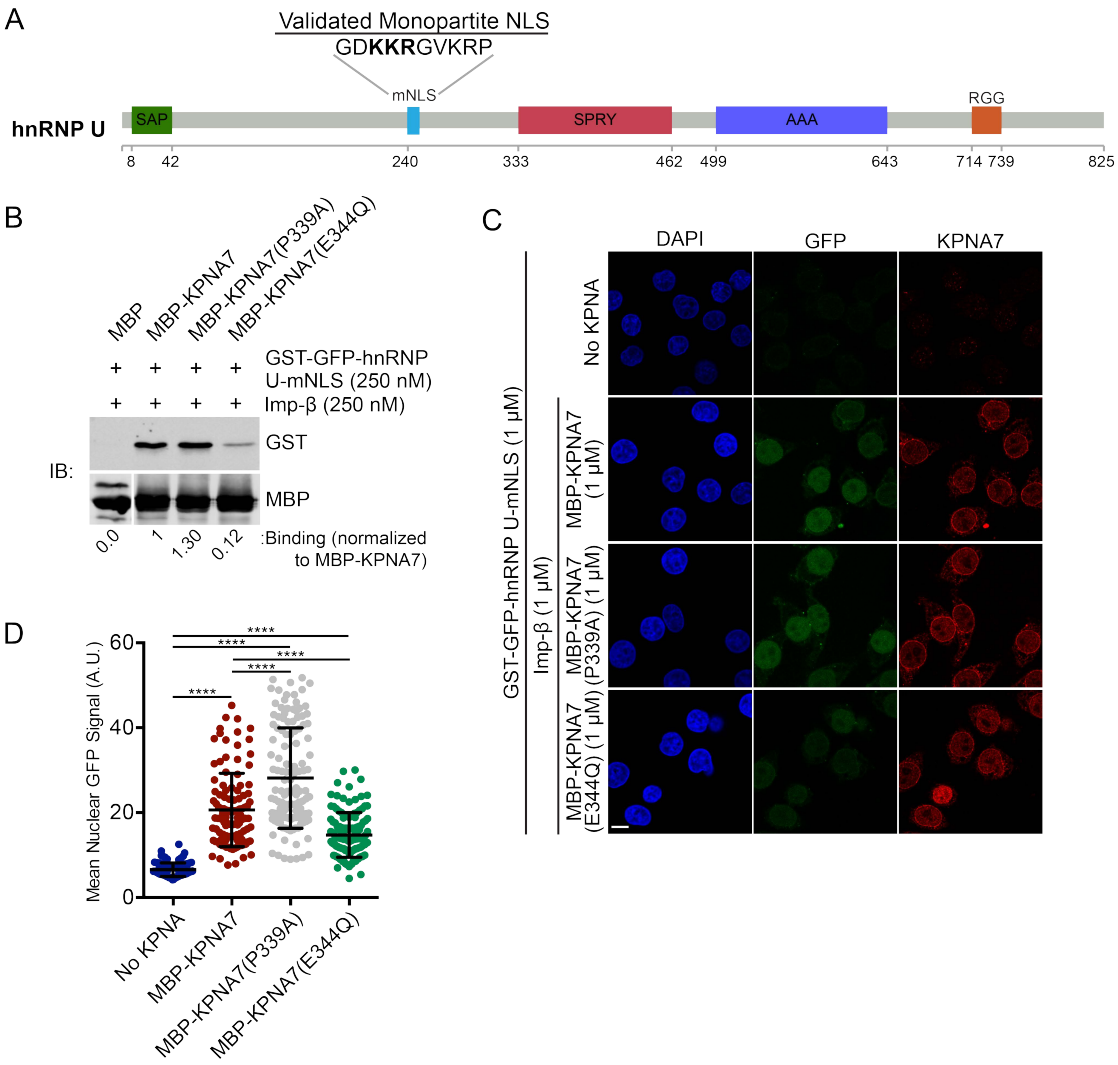


Figure 3.6: KPNA7 binding to the monopartite NLS of hnRNP U is reduced by the E344Q substitution.
(A) Representation of hnRNP U protein. Domains shown are as follows: SAP – SAP DNA binding motif; mNLS – monopartite NLS sequence, SPRY – SPRY domain unknown fucntion,; AAA – ATPase domain; RGG – RNA-binding Arg-Arg-Gly box. **(B)** The monopartite NLS sequence from hnRNPU was expressed as a GST-GFP fusion protein and used for a binding assay with MBP-KPNA7 proteins on amylose beads in the presence of Imp-β (250 nM). Normalized binding was calculated and is shown below each lane. **(C)** Fluorescence microscopy images of permeabilized cell import assays with GST-GFP-hnRNP U NLS. The assays were performed in HeLa cells permeabilized with 0.005% digitonin, and incubated with Ran (0.2 μM), NTF2 (2 μM) and an energy regenerating system with the recombinant factors indicated. Confocal IF microscopy was used to visualize. **(D)** Quantification of NLS import for each reaction. At least 100 cells of each were quantified. (Scale: 10 μm; ****p<0.0001)

I. *KPNA7 expression is induced during neurogenesis*

The effect of mutations in KPNA7 on neuronal development suggests a function for KPNA7 in this process. To determine if KPNA7 undergoes expression changes during human neurogenesis we utilized human induced pluripotent stem cell (iPSC)-derived neural progenitor cells (NPCs) and observed expression fold-changes of the seven Imp-α isoforms over seven weeks of neurogenic factor-induced differentiation (**Fig 3.7A**). We observe a significant increase (~2.3-fold) in *KPNA7* expression from NPC (0 weeks differentiation) to three-week-old neurons. After seven weeks of differentiation, expression has increased ~3.6-fold relative to NPCs. It should be noted however that at seven weeks of differentiation *KPNA7* transcript levels are still at least ~15-fold lower than the other Imp-α isoforms. Regarding the other isoforms, *KPNA2* is highly expressed in NPCs but expression reduces significantly (~6.2-fold) during differentiation. Additionally, there is a small increase (~0.7-fold) in the expression of *KPNA5* by three weeks of differentiation. There are minimal changes in expression of the other Imp-α isoforms. From these data, we conclude that there are changes in expression levels of Imp-α isoforms during neuronal differentiation, including induction of *KPNA7* expression which suggests a role for KPNA7 function in neurons.

J. *KPNA7 is highly expressed in the developing human embryo*

Previous studies in mice, cattle, pigs, and rainbow trout have demonstrated highly restrictive expression of *KPNA7* homologs, primarily to the ovary, oocyte, and early developing embryo (Hu et al., 2010; Tejomurtula et al., 2009; Wang et al., 2014, 2012). In mice, cattle, and pigs, disruption of *KPNA7* expression in embryogenesis results in embryos which fail to progress to the blastocyst stage (Hu et al., 2010; Tejomurtula et al., 2009; Wang et al., 2012). We evaluated the expression of *KPNA7* during human embryogenesis by utilizing a single-cell RNA-seq dataset of human preimplantation embryos from Yan and colleagues (Yan et al., 2013) (**Fig 3.7B**). *KPNA7* is the highest expressed Imp-α isoform in the oocyte, zygote, 2-cell embryo, and 4-cell embryo (greater than 1000 RPKM in each). Expression levels decrease precipitously in the 8-cell embryo and morulae, and expression is nearly absent in the late blastocyst. Additional interesting expression changes are observed with large increases in expression of *KPNA2* and *KPNA6* from oocyte to zygote. From these data, we conclude that high *KPNA7*

expression in the developing embryo is conserved in humans and a rapid decrease in expression is unique to *KPNA7* among the Imp-α isoforms. These conclusions suggest a strong regulatory mechanism for the *KPNA7* gene.

K. *Exon 7 of KPNA7 contains a CTCF binding motif*

The basis of *KPNA7* expression change during development is not defined. We noted, however, that exon 7 of *KPNA7* contains a potential binding site for the transcription factor CCCTC-binding factor (CTCF) (c.1020-1038). CTCF is a transcriptional insulator made up of 11 zinc-finger DNA-binding domains (Merkenschlager and Nora, 2016). To determine if CTCF is bound to exon 7 of *KPNA7* in stem cells, neural progenitors cells and neurons we utilized CTCF ChIP-Seq data from the Encode Project (ENCFF619IWL, ENCFF940XMP, ENCFF259PXQ) (**Fig 3.8A**). In hESCs, there is CTCF signal at exon 7 of *KPNA7* (red box), while this signal is greatly reduced in an hESC-derived neural progenitor cell and appears absent in the hESC-derived neural cell. It should be noted that these decreases also correspond with a general global decrease in CTCF signal which occurs during neuronal differentiation (Beagan et al., 2017) and can also be seen in the CTCF peak downstream of *KPNA7*. To evaluate if CTCF may be a general regulator of Imp-α isoform switching in neurogenesis we examined the other 6 Imp-α genes in the same CTCF Chip-Seq datasets (**Fig 3.8B**). Changes in intronic CTCF peaks are observed in *KPNA1*, *KPNA3*, and *KPNA6* (black boxes), however, none of these peaks track with differentiation, or expression changes in each gene, in a similar manner to the peak in exon 7 of *KPNA7*.

The CTCF binding motif in exon 7 (c.1020-1038) of *KPNA7* overlays the epilepsy-associated c.1030G>C transversion, the causal mutation of the E344Q substitution (**Fig 3.9A**). The CTCF binding sequence logo shows an equal frequency of guanine and adenine at position 11. The transversion replaces guanine with cytosine and is predicted to disrupt CTCF binding. To evaluate the effect of this mutation on the affinity of CTCF for the *KPNA7* sequence, we performed polarization anisotropy measurements with a recombinant portion of CTCF (Zinc fingers 4-8) and fluorescently labeled dsDNA oligos of the wild-type and mutant KPNA7 sequences (Fig 9B). CTCF bound to the wild-type sequence (apparent $K_D = 1.76 \pm 0.084$ μM), while the mutant sequence bound at a

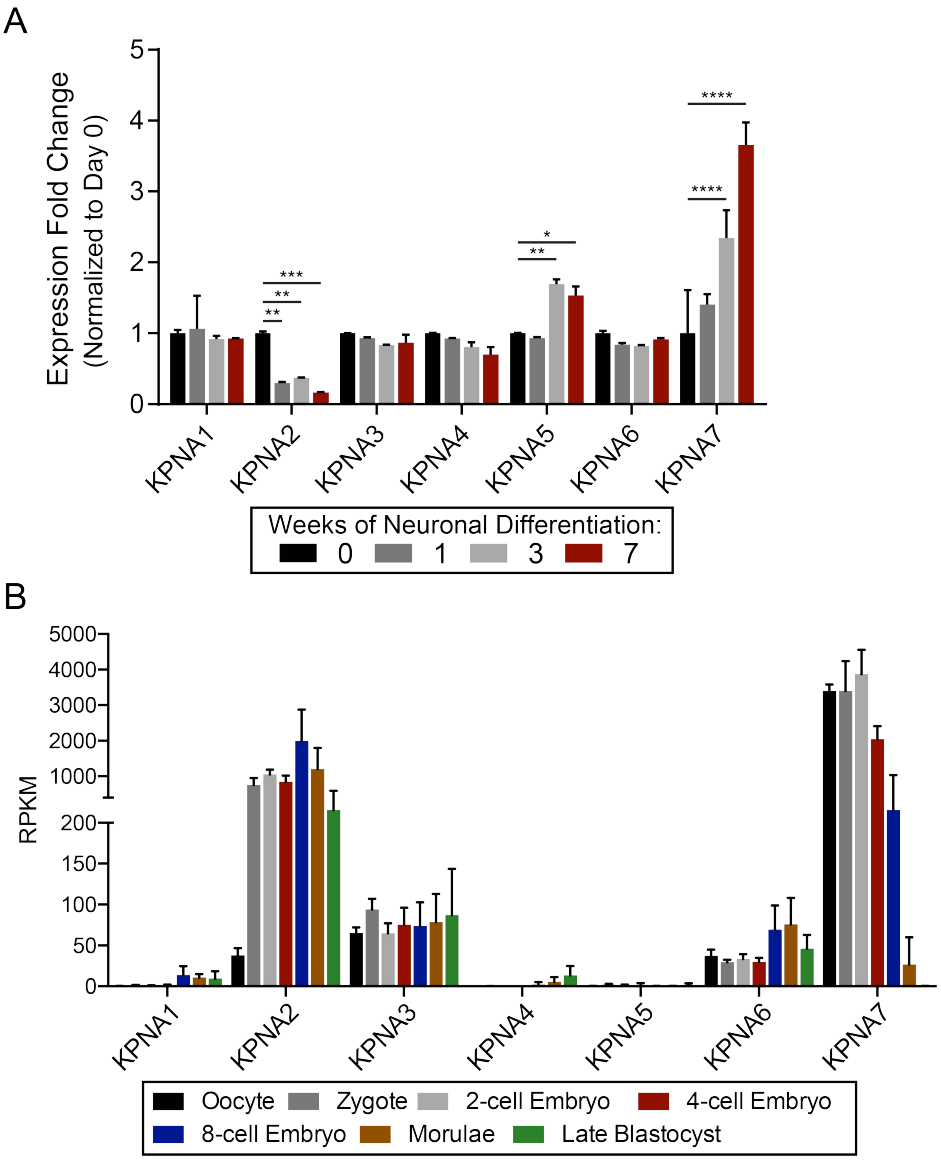
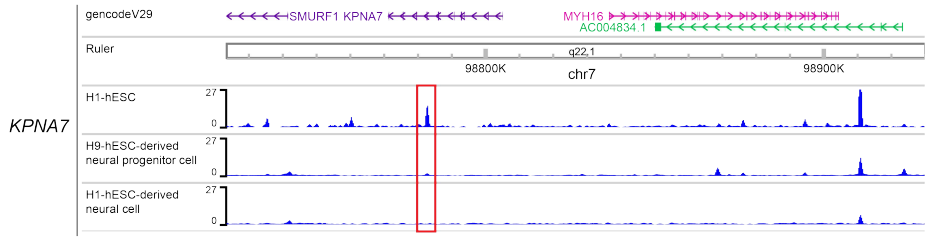


Figure 3.7: KPNA7 expression levels change during human neuronal differentiation and embryo development.
(A) RT-qPCR analysis of the human Imp-α genes during the neuronal differentiation of iPSC-derived neural progenitor cells (NPCs). Expression is normalized to the NPCs at 0 weeks of differentiation for each isoform. Significant differences in expression are indicated. (****p<0.0001, ***P<0.001, **p<0.01, *p<0.05) **(B)** Expression levels of the Imp-α genes in single-cell RNA-seq dataset from Yan and colleagues (Yan et al., 2013) for the indicated human cell types.

significantly lower affinity (apparent K_D = 76.04 ± 16.20 μM) (p=0.0014). This suggests that CTCF binding to exon 7 of KPNA7 would be significantly reduced by the c.1030 G>C transversion.

Finally, we evaluated the effect of loss of CTCF on expression of *KPNA7* in NPCs. We generated stable NPC lines with shRNAs targeting CTCF and a scrambled control to evaluate the effect of CTCF knockdown on *KPNA7* expression in these cells (**Fig 3.9A, B**). While the control shRNA had no effect on *KPNA7* expression levels, there was a significant increase in expression with knockdown of CTCF (~3.3-fold, p<0.001). This result suggests that CTCF plays a role in regulating *KPNA7* expression in NPCs. On the basis of these data, combined with the ChIP-Seq analysis, CTCF binding to exon 7 of *KPNA7* is predicted to repress mRNA expression and loss of this binding may play a role in the induction of *KPNA7* during neurogenesis. Finally, the c.1030G>C is predicted to disrupt this interaction and may result in dysregulation of *KPNA7* expression.

A


KPNA7
gencodeV29
SMURF1 KPNA7
MYH16
AC004834.1
Ruler
q22.1
98800K
chr7
98900K
H1-hESC
H9-hESC-derived neural progenitor cell
H1-hESC-derived neural cell
27
0

B

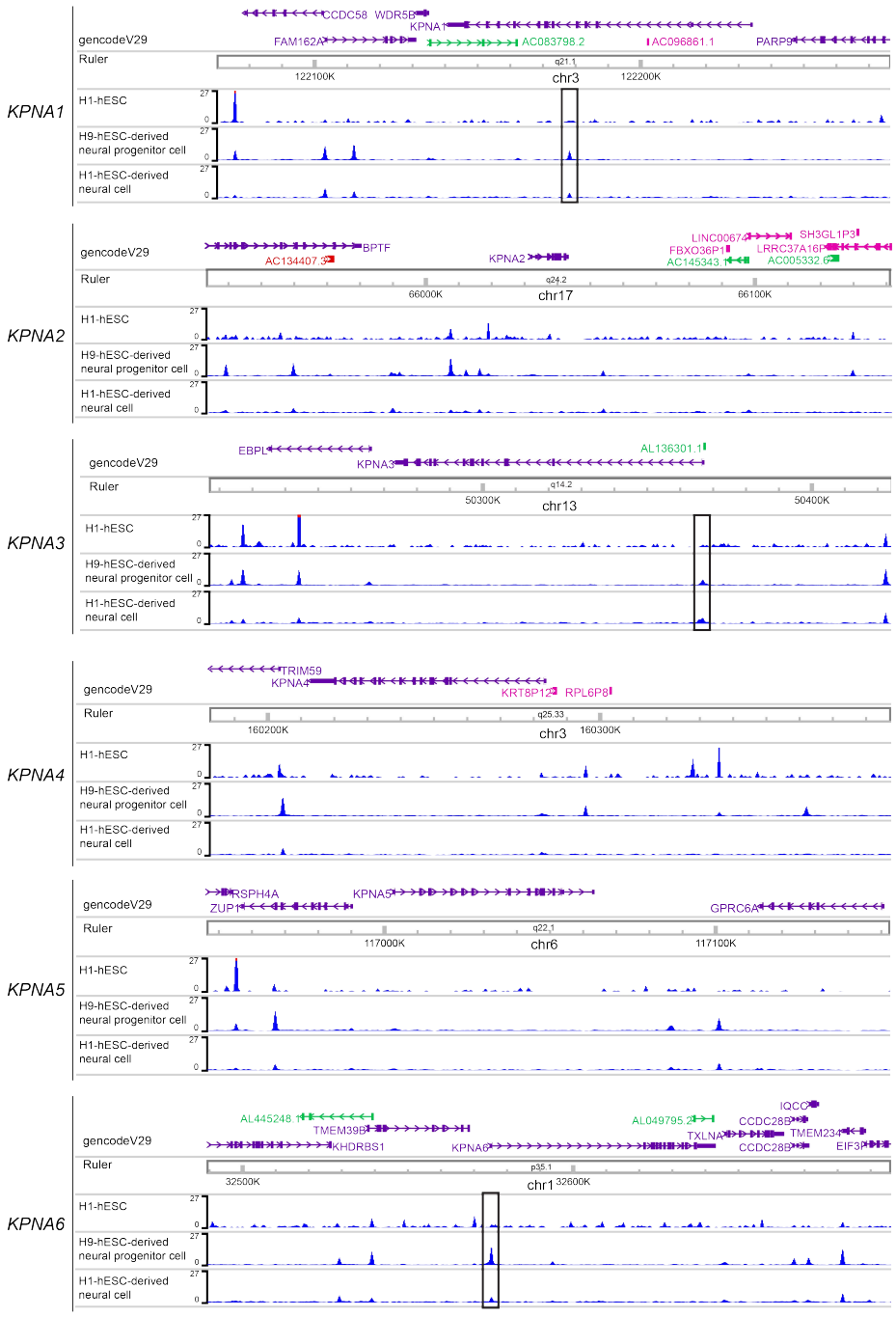

KPNA1
CCDC58 WDR5B
KPNA1
gencodeV29
FAM162A
AC083798.2
AC096861.1
PARP9
Ruler
q21.1
122100K
chr3
122200K
H1-hESC
H9-hESC-derived neural progenitor cell
H1-hESC-derived neural cell
KPNA2
BPTF
AC134407.3
KPNA2
LINC00674
SH3GL1P3
FBXO36P1
LRRC37A16P
AC145343.1
AC005332.6
q24.2
66000K
chr17
66100K
KPNA3
EBPL
KPNA3
AL136301.1
q14.2
50300K
chr13
50400K
KPNA4
TRIM59
KPNA4
KRT8P12
RPL6P8
q25.33
160200K
chr3
160300K
KPNA5
RSPH4A
ZUP1
KPNA5
GPRC6A
q22.1
117000K
chr6
117100K
KPNA6
AL445248.1
TMEM39B
KHDRBS1
KPNA6
AL049795.2
IQCC
CCDC28B
TXLNA
TMEM234
CCDC28B
EIF3I
p35.1
32500K
chr1
32600K

Figure 3.8: CTCF occupancy of *KPNA7* exon 7 is lost during neuronal differentiation of stem cells.
(A) CTCF ChIP-Seq profiles from three ENCODE datasets (ENCFF619IWL, ENCFF940XMP, ENCFF259PXQ) of H1-ESCs, H9-ESC-derived neural progenitor cell, and an H1-ESC-derived neural cell at the *KPNA7* gene locus. A red box indicates the CTCF peak on Exon 7 of *KPNA7* which is lost during differentiation. **(B)** CTCF ChIP-Seq profiles from the same datasets for the other six Imp-α genes. Black boxes are shown around peaks in each gene which changes with differentiation.

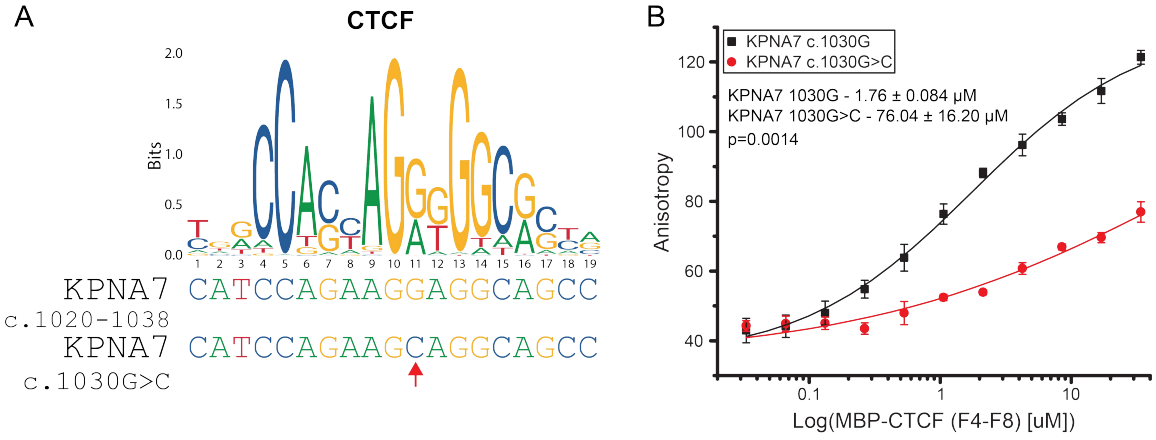


Figure 3.9: The 1030G>C transversion significantly reduces CTCF binding affinity.
(A) CTCF binding motif with the *KPNA7* sequences shown underneath. The c.1030G>C transversion is marked with a red arrow. **(B)** Fluorescence polarization anisotropy analysis of recombinant MBP-CTCF (Zinc Fingers 4-8) binding to fluorescently labeled dsDNA oligos of the 16 base-pair *KPNA7* CTCF binding motif. The G>C transversion reduces CTCF affinity from 1.76 ± 0.084 µM to 76.04 ± 16.2 µM.

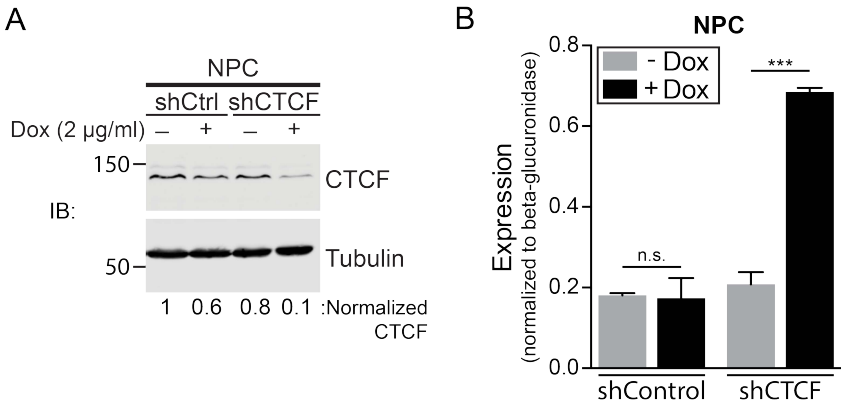


Figure 3.10: CTCF knockdown in neural progenitor cells increases *KPNA7* expression. **(A)** Western blot analysis of CTCF protein levels in shCTCF and shCtrl NPCs with and without doxycycline treatment for 7 days. CTCF protein levels were normalized to tubulin and then relative to the control condition. **(B)** RT-qPCR expression analysis of *KPNA7* transcript levels in shCTCF and shCtrl NPCs. Expression was normalized to beta-glucuronidase. Significance is shown. (****p<0.0001)

IV. Discussion

The Imp-α family of transport receptors share between 40 and 85% identity but have a highly conserved structure which consists of an auto-inhibitory N-terminal IBB domain and a body made up of ARM repeats (Conti et al., 1998; Kelley et al., 2010). Each ARM repeat consists of about 40 amino acids that make up a three alpha-helix secondary structure (Peifer et al., 1994). ARM-repeats, which are also in unrelated proteins such as β-catenin, stack to form a supercoiled tertiary structure. In Imp-α, the stacking of 10 ARM repeats creates a concave surface used for NLS binding (Conti et al., 1998).

The Imp-α NLS binding surface is separated into a major NLS groove toward the N-terminus of the receptor, between ARMs 2-4, and a minor NLS groove which is toward the C-terminus, between ARMs 6-8. The major NLS-binding groove is responsible for binding both monopartite and bipartite NLS sequences, while the minor groove is primarily used in binding to the second element of the bipartite NLS. The two epilepsy-associated amino acid substitutions in KPNA7 map to ARM repeat 7 in the minor NLS binding groove (**Fig 3.1A**). This would suggest that if the mutants were to impact NLS binding by the receptor it would primarily be in the context of binding bipartite NLS sequences. Gln344 is conserved in all seven human Imp-α isoforms (**Fig S3.3A**). Substitution at this position of an uncharged residue, Gln for the negatively charged Glu, could be predicted to impact binding of the positively charged NLS sequence. Analysis of the structure of KPNA2 bound to the bipartite Rb NLS (Fontes et al., 2000) shows that the homologous E354 residue does not, in fact, directly interact with the NLS, but hydrogen bonds with a conserved Arg in ARM 6 (**Fig S3.3B**). It is this Arg (315 in KPNA2) which directly interacts with the bound NLS and E344 stabilizes this interaction. In addition to reducing the binding of bipartite NLS sequences directly, disruption of this inter-ARM interaction might alter the overall structure of the ARM scaffold to impact NLS binding in the major groove as well. This is a potential explanation for our observations that KPNA7(E344Q) has reduced function toward both monopartite NLS sequences (from SV40 and hnRNP U) and bipartite NLS sequences (from DDB2 and hnRNP R).

The identity of the amino acid residues which separate the first and second helix of each ARM repeat has been shown to impact the overall flexibility of Imp-α (Pumroy et al., 2015). Across many unrelated ARM containing proteins, there is conservation of glycine or proline residues in ARM repeats which help to facilitate the turn between helices (Andrade et al., 2001). The proline residue at position 339 of KPNA7 is not directly conserved, however, the other six human isoforms do have a conserved proline in this loop between the second and third helices of ARM 7 (**Fig S3.3A**). The shift in the position of this proline residue places it closer to helix 3 of ARM7 in KPNA7 versus KPNA2 (**Fig S3.3C, D**). In both cases, however, the rigid dihedral angle of proline is expected to help facilitate the turn between helices two and three. The substitution of alanine at this position may impact the structure of ARM 7 but based on our data does not result in significant changes in NLS binding or transport.

Modulation of nuclear import via differential expression of Imp-α isoforms is critical for multiple developmental programs (Geles and Adam, 2001; Köhler et al., 1997; Mason et al., 2002; Mihalas et al., 2015). A process termed Imp-α isoform switching has been described in mouse neuronal differentiation (Yasuhara et al., 2006). In mice, differential expression of Kpna1 and Kpna2 results in distinct transport outcomes for pro-pluripotency and pro-neural NLS containing proteins including Oct3/4, Sox2, Oct6 and Brn2 (Yasuhara et al., 2013). These changes function to promote either a stem-like or neuronal state. We observed Imp-α isoform switching during the neuronal differentiation of human iPSC-derived NPCs to neurons. In agreement with the mouse data, *KPNA2* expression decreased during differentiation. In contrast, *KPNA1* expression remained relatively constant which may reflect differences in regulation of isoforms between human and mice. Expression of *KPNA7* increased significantly during differentiation. This result, together with the severe neurodevelopmental defect associated with mutation in KPNA7 which reduces protein function, prompted us to interrogate KPNA7 interacting proteins in neurons.

We utilized human iPSC-derived neurons and SILAC-based MS to identify KPNA7 interacting proteins. This was the first study to query KPNA7 interactions in neurons and adds to previous reports of KPNA7 binding partners in HEK293 and pancreatic cancer

cell lines (Kimoto et al., 2015; Vuorinen et al., 2017). In agreement with these studies, we identified significant enrichment in proteins involved in RNA processing among the KPNA7 binding partners. We validated the interaction of KPNA7 with two hnRNP proteins, hnRNP R and hnRNP U, both in the context of binding to the full-length proteins from neurons and to recombinant NLS sequences from each protein. In the course of our analysis, we also determined the functionality of previously described, putative mono- and bipartite NLS sequences in hnRNP R (Cappelli et al., 2018; Mourelatos et al., 2001). Nuclear import of hnRNP R was found to be primarily dependent on the presence of the bipartite NLS sequence near the C-terminus of the protein. Additionally, multiple Imp-α isoforms bound the hnRNP R-bNLS while no binding was observed to the putative mNLS sequence. In regard to KPNA7 binding to hnRNP R and hnRNP U, we observed a reduction in binding by KPNA7(E334Q) to each and further measured reduced transport function for the KPNA7 mutant toward each functional NLS sequence. This reduced transport function for the E344Q mutant neuronal KPNA7 interacting proteins suggests a potential mechanism for the negative impact on neuronal development associated with KPNA7 mutation.

Both hnRNP R and hnRNP U have been linked to neuronal function and development. hnRNP R associates with the survival motoneuron protein (SMN) in axons to promote neurite outgrowth in neuronal differentiation (Dombert et al., 2014; Rossoll et al., 2003). Mutations in SMN are responsible for the severe developmental motoneuron disorder spinal muscular atrophy (Lefebvre et al., 1995), and loss of hnRNP R reduces axonal growth similarly to loss of SMN (Dombert et al., 2014; Glinka et al., 2010). Additionally, mutations in hnRNP R, including one in the bipartite NLS, have been associated with severe developmental delay, seizures, and corpus callosum and cerebellar abnormalities (Duijkers et al., 2019). Variants in hnRNP U have been linked to neuronal development and disorders such as encephalopathies, intellectual disabilities, and epilepsy (Bramswig et al., 2017; Carvill et al., 2013; Consortium et al., 2013; Hamdan et al., 2014). A study in *C. elegans* identified a potential mechanism by demonstrating regulation of expression of the potassium channel Slo2 by hnRNP U (Liu et al., 2018). These studies suggest that disrupted regulation of hnRNP R and hnRNP U could result in severe

neurodevelopmental defects similar to those observed in the patients with mutations in KPNA7.

Finally, we identified potential CTCF dependent regulation of KPNA7 expression via CTCF binding to exon 7 of the *KPNA7* gene. CTCF has multiple functions but primarily serves to form the border between chromatin topological associating domains (TADs). Across the human genome, around 40% of CTCF binding sites are intragenic with 5% in exons (Chen et al., 2012). Specific roles for exonic CTCF binding have been determined and include transcription inhibition (Renaud et al., 2005) and regulation of alternative splicing (Shukla et al., 2011). Previous studies have shown both a necessity for CTCF in neuronal development, as well as a general decrease in both *CTCF* expression and the number of CTCF ChIP-Seq peaks during neuronal differentiation (Beagan et al., 2017; Hirayama et al., 2012; Sams et al., 2016). We suggest that a reduction in CTCF binding to KPNA7 during neuronal differentiation plays a role in inducing *KPNA7* expression. Additionally, the c.1030G>C transversion identified in patients with severe neurodevelopmental defect significantly reduces CTCF affinity for the *KPNA7* binding motif. It also remains possible that the other epilepsy-associated mutation, c.1015C>G, might also reduce CTCF binding despite residing outside of the core CTCF binding motif. Furthermore, while we did not directly interrogate whether CTCF functions to reduced CTCF binding to ARM 7 of *KPNA7* might function to

In conclusion, we have characterized an epilepsy-associated mutation (c.1030G>C) in KPNA7 which is predicted to reduce binding of a regulatory protein (CTCF) to the *KPNA7* gene and results in reduced NLS binding and transport function of the translated protein (KPNA7(E344Q)) for both mono- and bipartite NLS sequences. Additionally, we have identified induction of expression of *KPNA7* in neurogenesis, and discovered two neuronal KPNA7 interacting proteins, hnRNP R and hnRNP U, and speculate that disruption of KPNA7 activity toward these two proteins might play a role in neurodevelopmental disorder.

V. Materials and Methods

A. *Protein Modeling*

The crystal structure of KPNA7 has not been solved. To generate a model structure for KPNA7 we utilized the SWISS-MODEL software to thread the KPNA7 primary sequence onto the crystal structure from mouse KPNA2 (PDB: 1PJM). The epilepsy-associated amino acid substitutions were made in PyMol. Images were rendered in PyMol (Schrodinger LLC, 2015).

B. *Plasmids*

The human KPNA7 protein sequence was codon optimized for expression in *E. coli* (Genewiz) and cloned into pMBPHis-Parallel1 (Sheffield et al., 1999) for expression as a maltose binding protein fusion. PCR mutagenesis was performed using Pfu Ultra II (Agilent) to introduce the P339A and E344Q amino acid substitutions. The NLS sequences used in this study (SV40, DDB2, hnRNP R mNLS, hnRNP R bNLS, hnRNP U) were cloned into the pGEX-4T1-GFP vector. Flag-myc-hnRNP R plasmid (RC224502) was obtained from OriGene. Mutagenesis PCR was performed using Pfu Ultra II (Agilent) to generate the NLS mutant constructs. The CTCF expression construct is in the pMAL-c2g backbone, which contains Zinc fingers 4-8 of human CTCF was a gift from Gary Felsenfeld and previously published (Renda et al., 2007).

C. *Recombinant Protein Expression*

Standard methods were utilized to express MBP-KPNA7 fusion proteins in BL21 *E. coli* at 18°C overnight. Cultures were harvested, lysed by French Press, clarified and incubated with amylose resin (NEB). After extensive washing, MBP-fusion proteins were eluted with 10 mM maltose and dialyzed into 50 mM Tris pH 8.0, 200 mM NaCl with 1 mM DTT. Standard methods were used to express each GST construct in BL21 *E. coli* at 18°C overnight, processed in the same manner as the MBP constructs and incubated with glutathione resin (Millipore-Sigma). GST-GFP-NLS constructs were dialyzed into PBS with 1 mM DTT. MCP-CTCF(F4-F8) was transformed into BL21 *E. coli* and expression was induced with 1 mM IPTG in the presence of 100 μM $ZnCl_2$ for 2.5 hours at 37°C. Cultures were harvested, and lysed by French press in 50 mM Tris pH 8, 200 mM KCl, 100 μM $ZnCl_2$, and 1 mM DTT with protease inhibitors. During lysis, Pierce Universal

Nuclease was included to digest nucleic acid bound to CTCF. The lysate was incubated with amylose resin and washed extensively with 2 M KCl. MBP-CTCF(F4-F8) was eluted with 10 mM Maltose in 20 mM Tris pH 8, 200 mM KCl, 100 µM $ZnCl_2$, 1 mM EDTA, 1 mM DTT, and 1 µg/ml each of leupeptin, pepstatin, and aprotinin. Purified protein was dialyzed into the same buffer to remove maltose. All recombinant proteins were analyzed by SDS-PAGE for purity and quantified Bradford protein assay. Methods for expression and purification of Imp-β, Ran, NTF2, and CAS have been published (Chatterjee and Paschal, 2015; Kelley et al., 2010; Kutay et al., 1997).

D. *Binding Assays*

MBP fusion proteins were bound to amylose resin for at least 3 hours at 4°C at 1 µg/µl before being blocked for at least 1 hour with 1 mg/ml BSA. Imp-β and GST-GFP-NLS constructs were incubated with 10 µl bead volume of the indicated resin at the indicated concentrations in 100 µl of total reaction volume for 2 hours at 4°C in amylose column buffer (20 mM Tris pH 8, 200 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.05% Triton X-100 and 1 µg/ml each of leupeptin, pepstatin, and aprotinin. After binding, beads were washed 5 times with 400 ul of the same buffer and analyzed by SDS-PAGE and western blot. Quantification for all blots was performed with ImageStudio (Li-Cor).

For the radiolabeled binding assay, the Imp-α isoforms were expressed as ^{35}S-methionine-labeled proteins by *in vitro* transcription/translation (TnT coupled reticulocyte lysate system; Promega). Scintillation counting was used to determine labeling incorporation and expression efficiency. GST-fusion proteins were bound to glutathione resin (1 µg/µl resin). Reactions were again 100 µl total volume and contained 10 µl of beads. 125,000 cpm of each labeled Imp-α protein was added to each reaction in 25 mM Tris pH 7.5, 50 mM NaCl, 5 mM MgCl2, 0.1 mM EDTA, 0.5 mg/ml BSA, 0.1% NP-40, and 1 mM DTT with leupeptin, pepstatin, and aprotinin (1 µg/ml each). Reactions were incubated for 3 hours at 4°C. After reactions, beads were washed 5X in the same buffer, analyzed by SDS-PAGE and developed on film.

E. *Cell Culture*

Neural progenitor cells derived from 9429 and BJ human fibroblasts (Coriell Cell Repository) were reprogrammed into induced pluripotent stem cells (iPSC) and induced

to neural progenitor cells as previously described (Brennand et al., 2011; Michel et al., 2019). NPC's were grown on Matrigel (Corning) in NPC medium, (DMEM/F12 + Glutamax (Invitrogen), 1X N2 (Invitrogen), 1X B27-Vitamin A (Invitrogen), 1ug/ml Laminin (Invitrogen), 20 ng/ml FGF-2 (Peprotech)), and passaged 1:6 every 4 days using Accutase (Innovative Cell Technologies). iPSC-derived neurons were cultured in neuron medium (DMEM/F12 + Glutamax, 1X N2, 1X B27 with Vitamin A (Invitrogen), 20 ng/ml BDNF (Shenandoah Biotechnology), 20 ng/ml GDNF (Shenandoah Biotechnology), 1 mM dibutyryl-cyclic AMP (Sigma), and 200 nM ascorbic acid (Stem Cell Technologies). Low passage (p4) BJ NPCs were used to generate doxycycline-inducible shRNA lines with control or CTCF targeted shRNAs. Transduced cells were selected by puromycin selection (0.4 μg/ml) for 72 hours and maintained in 0.25 μg/ml puromycin. For knockdown experiments, shRNA expression was induced with 2 μg/ml doxycycline for 7 days.

Adherent HeLa cells (ATCC) were grown in DMEM-High Glucose (Gibco) with 10% FBS (Atlanta Biologics). HeLa cells were transfected with ViaFect reagent (Promega). HEK293T cells (ATCC) were cultured in in DMEM (Gibco) with 5% FBS (Atlanta Biologics), 1% non-essential amino acids (Gibco), 1% sodium pyruvate (Gibco) and 1% Pen/Strep (Gibco). The HA-KPNA7 HEK293T cell line was described in Chapter II of this study. Lines expressing HA-KPNA7(P339A) and HA-KPNA7(E344Q) were similarly generated.HEK293T cells were transfected with Fugene6 reagent (Promega).

F. *Permeabilized cell import and export assays*

Digitonin-permeabilized cell import assays were performed as previously described (Cassany and Gerace, 2009), and as described in Chapter II of this text. In this current chapter, HeLa cells were plated in coverslips at a density of $0.15x10^6$ cells per ml 24 hours before each assay. Cells were washed 3x with ice-cold transport buffer (20 mM HEPES [pH 7.4], 110 mM potassium acetate, 2 mM magnesium acetate, 0.5 mM EGTA) containing 1 mM dithiothreitol (DTT) and 1 μg/ml portions (each) of leupeptin, pepstatin, and aprotinin and then permeabilized with 0.005% digitonin for 5 min on ice. Import reactions contained an energy-regenerating system (5 mg/ml of bovine serum albumin, 80 U of creatine phosphokinase/ml, 1.6 mg of creatine phosphate/ml, 1 mM ATP, 1 mM

GTP), 1 μM NTF2 and 2 μM Ran, as well as 1 μM import substrate (GST-GFP fusion), 1 μM Importin-β and/or 1 μM MBP-KPNA7 as indicated. Reactions were carried out for 30 minutes at 30°C and terminated by transferring to ice-cold transport buffer. After 3 additional washes in transport buffer, coverslips were processed for confocal immunofluorescence microscopy as described below. Quantification of mean nuclear intensity was quantified using ImageJ software by making a mask with DAPI channel. One-way ANOVA with multiple comparisons was used for statistical analysis in GraphPad Prism. Export assays were performed as described in Chapter II of this study using wild-type and mutant HA-KPNA7 cell lines.

G. *2-Color bead-based fluorescence imaging assay*

The bead-based fluorescence imaging assay used in this chapter was adapted from (Patel and Rexach, 2008) and used in Chapter II of this current text. Briefly, amylose resin was pre-loaded with protein (1 μg/μl packed beads) for at least 4 hours at 4°C in 20 mM Tris pH 8, 200 mM NaCl, 1 mM EDTA, 1 mM DTT with leupeptin, pepstatin, and aprotinin (1 μg/ml each) and blocked with 5 mg/ml BSA. MBP was labeled with IRDye-680 (LiCor) and added to control MBP beads at 0.05 μg/μl. MBP and MBP-KPNA7 beads were then combined at a ratio of 1:1 and made to a 50% slurry. Reactions were assembled in siliconized microcentrifuge tubes with 1.5 μl bead slurry, 1 μl 4X reaction buffer (40 mM EDTA, 40 mg/ml BSA, 500 mM NaCl), x μl GST-GFP cargo (750 nM, final concentration), x μl Imp-β (750 nM, final concentration), and distilled water to a total volume of 4 μl. Reactions were then incubated for 10 minutes at room temperature, washed 2x with 200 μl of 1X reaction buffer and transferred in 20 μl total volume to a 96 well plate well for imaging. Beads were imaged by phase contrast, and with a Cy5 and GFP filter using the EVOS Fl Imaging System (Invitrogen) with a 20x objective at equivalent intensity and exposures. The GFP and 680 signals were quantified using ImageJ and separated into 680 positive and negative (MBP and MBP-KPNA7 beads respectively) and GFP signal was plotted. GraphPad Prism was used for statistical analyses.

H. *KPNA7 cargo identification by SILAC based-Mass Spectrometry*

9429 NPCs were plated onto Matrigel in T75 culture flasks. At 80% confluency neuronal differentiation was induced by switching to neuron media. Media was changed every 4 days for 2 weeks. Plates were gently washed 2x with PBS and switched to SILAC neuron medium. Basal SILAC neuron media contains DMEM/F12 for SILAC (ThermoFisher), which is deficient in Lysine and Arginine, 1X Glutamax (ThermoFisher), and the supplements in neuron media described above. Heavy SILAC media was made by supplementing the basal medium with 0.669 mM L-Arginine:HCl ($^{13}C_6$, $^{15}N_4$) (Cambridge Isotope Laboratories) and 0.498 mM L-Lysine:HCl ($^{13}C_6$, $^{15}N_2$) (Cambridge Isotope Laboratories). Light SILAC media was made by supplementing the basal medium with 0.669 mM L-Arginine:HCl (ThermoFisher) and 0.498 mM L-Lysine:HCl (ThermoFisher). Neurons were grown for 2 more weeks in each SILAC medium with media changes every three days for a total of 4 weeks of differentiation. Cells were harvested by scraping into cold PBS with 1 mM PMSF and pelleted by centrifugation. To check incorporation of the heavy amino acids, total cell lysate was analyzed by liquid chromatography-mass spectrometry (LC-MS). Greater than 1000 proteins were identified with an average heavy incorporation of ~85%.

Nuclear lysates were generated by incubating the neuronal pellets in 25 mM HEPES pH 7, 25 mM KCl, 0.05 mM EDTA, 5 mM $MgCl_2$, 0.1% NP-40, 10% glycerol, and 1 mM DTT with leupeptin, pepstatin, and aprotinin (1 μg/ml each) and incubating with shaking at 4°C for 15 minutes. The insoluble nuclei were pelleted by centrifugation at 3K rpm for 5 minutes. Nuclear proteins were then released by 10-minute incubation in 50 mM HEPES pH 7.6, 400 mM KCl 0.1 mM EDTA and 1 mM DTT with leupeptin, pepstatin, and aprotinin (1 μg/ml each). Lysates were then spun for 20 minutes at 13K rpm and the soluble nuclear lysate was removed for use in pull-downs. Lysates were quantified by Bradford protein assay. Pull-downs were performed by immobilizing 40 μg of MBP or MBP-KPNA7 on magnetic amylose resin (NEB). Beads were washed extensively with amylose column buffer. Equal amounts of the heavy and light lysates were loaded onto the MBP-KPNA7 and MBP beads respectively in the presence of 0.5 μM his-Importin β. Reactions were incubated with agitation at 4°C for 3 hours and then washed with amylose column buffer 5 times.

The light and heavy samples were combined in equal amounts for LC-MS/MS analysis performed by the W.M. Keck Biomedical Mass Spectrometry Laboratory at the University of Virginia. The sample was reduced with DTT (10 mM, 1 hr, RT) followed by alkylation with iodoacetamide (50 mM, 1 hr, RT). The sample was then digested overnight with alkylated trypsin (1 µg; Promega, V5111). The digestion was quenched with 3 µl acetic acid. The mixed solution was evaporated to 20 µL for MS analysis. The LC-MS system consisted of a Thermo Electron Q Exactive HFX mass spectrometer system with an Easy Spray ion source connected to a Thermo 75 µm x 15 cm C18 Easy Spray column. The sample was injected, and the peptides eluted from the column by an acetonitrile/0.1 M formic acid gradient at a flow rate of 0.3 µL/min over 2.0 hours. The nanospray ion source was operated at 1.9 kV. The digest was analyzed using the rapid switching capability of the instrument acquiring a full scan mass spectrum to determine peptide molecular weights followed by product ion spectra (10 HCD) to determine amino acid sequence in sequential scans. This mode of analysis produces approximately 30000 MS/MS spectra of ions ranging in abundance over several orders of magnitude. The data were analyzed by database searching using the Sequest search algorithm against Uniprot human. The abundance ratios of all identified proteins were generated using ProteomeDiscoverer software. There were 41 proteins with greater than 10 identified peptides and an abundance ratio of greater than 2.75. PANTHER (pantherdb.org) was used to analyze these 41 genes for gene ontology biological process enrichment.

I. *Confocal Immunofluorescence Microscopy*

Coverslips from transport and transfection assays were fixed with 3.75% formaldehyde for 15 minutes, permeabilized with 0.02% Triton-X for 5 minutes and blocked with 2% FBS and 2% BSA in PBS for 1 hour. Primary antibodies (anti-Flag M2 (mouse monoclonal, F1804, Sigma), anti-LaminA (rabbit polyclonal, L1293, Sigma), anti-KPNA7 (rabbit polyclonal, HPA031395, Sigma)) were diluted in blocking buffer and incubated at room temperature for 2 hours. Secondary antibodies (Goat Anti-Rabbit-Cy3 (115-165-144, Jackson ImmunoResearch), Donkey anti-mouse-FITC (715-225-151, Jackson ImmunoResearch)) were diluted in blocking buffer and incubated at room temperature for 1 hour. DAPI was used to stain nuclei. Images were acquired by laser

scanning confocal microscopy (Zeiss 880 LSM, Carl Zeiss) at 40×, 1.3 NA oil immersion objective with Zen software (Carl Zeiss).

J. *Co-Immunoprecipitation*

For co-immunoprecipitation experiments, cells were harvested in ice-cold PBS with 1 mM PMSF and pelleted. Pellets were lysed for 20 minutes on ice in a PBS buffer with 100 mM NaCl, 0.5% Triton, 2.5 mM EDTA, 2 mM DTT, and protease inhibitors, and then tip sonicated. After sonication lysates were clarified and loaded onto antibody beads (anti-HA agarose (Sigma) or protein G beads with hnRNP U primary antibody (Santa Cruz)) and incubated at 4°C for 4 hours with end-over-end turning. Samples were washed with PBS with 100 mM NaCl, 0.5% Triton, 1 mM EDTA, 2 mM DTT, and protease inhibitors and then analyzed by SDS-PAGE and western blot.

K. *Gene Expression Analysis*

RNA was harvested by Trizol extraction per manufacturer protocol. RT-qPCR was performed per standard methods using the following primer sequences:

KPNA1 (5’-TAGCAACATTTCTCCGCTTG-3’ and 5’TCTCTGAAT-CCCGATGAGATG-3’), KPNA2 (5’-TGATTTTCCACATTGCTGCT-3’ and 5’-GATGATGCTACTTCTCCGCTG-3’), KPNA3 (5’-TTTTGTTCTTCCGCAGTTCC-3’ and 5’-CGCATCAAGAGCTTCAAGAAC-3’), KPNA4 (5’-CAACTTCATTTCGTTGTCTTCTC-3’ and 5’-CGGACAACGAGAAACTGGAC-3’), KPNA5 (5’-CGGCATTTCTTGTTGTTGTG-3’ and 5’-TGCTGGTGACAATGCAGAAT-3’), KPNA 6 (5’-AATTGTCTTTCCCTGGGCTC-3’ and 5’-ATTGTCTACTGAAAGCTGCCG-3’), and KPNA 7 (5’-CATCGAGAAGCACTTTGGTG-3’ and 5’-GGAGGTAGGGAGCTTGGCTA-3’).

L. *Fluorescence polarization anisotropy*

Single strand forward and reverse DNA oligonucleotides of the KPNA7 exon 7 CTCF binding motif (c.1021-1036) with and without the c.1030G>C transversion were purchased with HPLC purification (IDT DNA Technologies). Fluorescein (6-FAM) was conjugated to the 5’ end of the reverse oligo for each pair. The sequences of the oligos are as follows: WT F: 5’-ATCCAGAAGGAGGCAG-3’; WT R: 5’-/56-FAM/CTGCCTCCTTCTGGAT-3’; Mutant F: 5’-ATCCAGAAGCAGGCAG-3’; Mutant

R; 5'-/56-FAM/CTGCCTGCTTCTGGAT-3'. Forward and reverse primers were annealed, and dsDNA was purified by ion exchange chromatography (QSepharose; GE Life Sciences). dsDNA was made up to10 nM in FP buffer (20 mM Tris pH 8, 200 mM KCl, 100 μM $ZnCl_2$, and 1 mM DTT) for fluorescence polarization (FP) experiments.

MBP-CTCF(F4-F8) was made up to 34.5 μM in FP buffer and titrated with two-fold serial dilutions for both the wild-type and mutant dsDNA oligos. Oligo concentration was kept constant at 10 nM. Anisotropy measurements were made in triplicate in a black COSTAR 96-well plate (Corning Life Sciences) on a PHERAstar microplate reader (BMG Labtech) with excitation at 494 nm and emission at 525 nm at 25°C. A binding curve was generated by fitting data to a sigmoidal modal using OriginPro Software. For the mutant oligo, which bound at a lower affinity, an apparent K_D was generated by setting the upper and lower bounds to the same as the wild-type oligo. An unpaired T-test was used to determine the significance of the difference in K_D.

M. *Data Sources*

Single-cell RNA-seq of human implantation embryos data from Yan and colleagues (Yan et al., 2013) was used to query the expression of the human *KPNA* genes. The Washington University Epigenome Browser (http://epigenomegateway.wustl.edu/browser/) was used to visualize CTCF Chip-Seq datasets from the ENCODE project (Sloan et al., 2016) with the following identifiers: ENCFF619IWL, ENCFF940XMP, ENCFF259PXQ.

N. *Antibodies*

Primary:

Anti-GST (B-14) – mouse monoclonal – Santa Cruz Biotechnology; sc-138

Anti-HA (16B12) – mouse monoclonal – Covance; MMS-101P

Anti-MBP – mouse monoclonal – NEB; E8032S

Anti-Imp-β (3E9) – mouse monoclonal – abcam; ab2811

Anti-KPNA7 – rabbit polyclonal – Millipore Sigma; HPA031395

Anti-hnRNP R – rabbit polyclonal – Millipore Sigma; SAB2700924

Anti-hnRNP U (3G6) – mouse monoclonal – Santa Cruz Biotechnology; sc-32315

Anti-Flag M2 – mouse monoclonal – Millipore Sigma; F1804

Anti-CTCF – rabbit monoclonal – Cell Signaling; D31H2

Acknowledgements: I would like to thank Sarah Goggin for taking an interest in my data after a McConnell lab meeting and noting the presence of the potential CTCF binding site in KPNA7. I would also like to thank Adam Boulton of the Bushweller Lab for his advice and assistance with the fluorescence polarization anisotropy experiment in this Chapter.

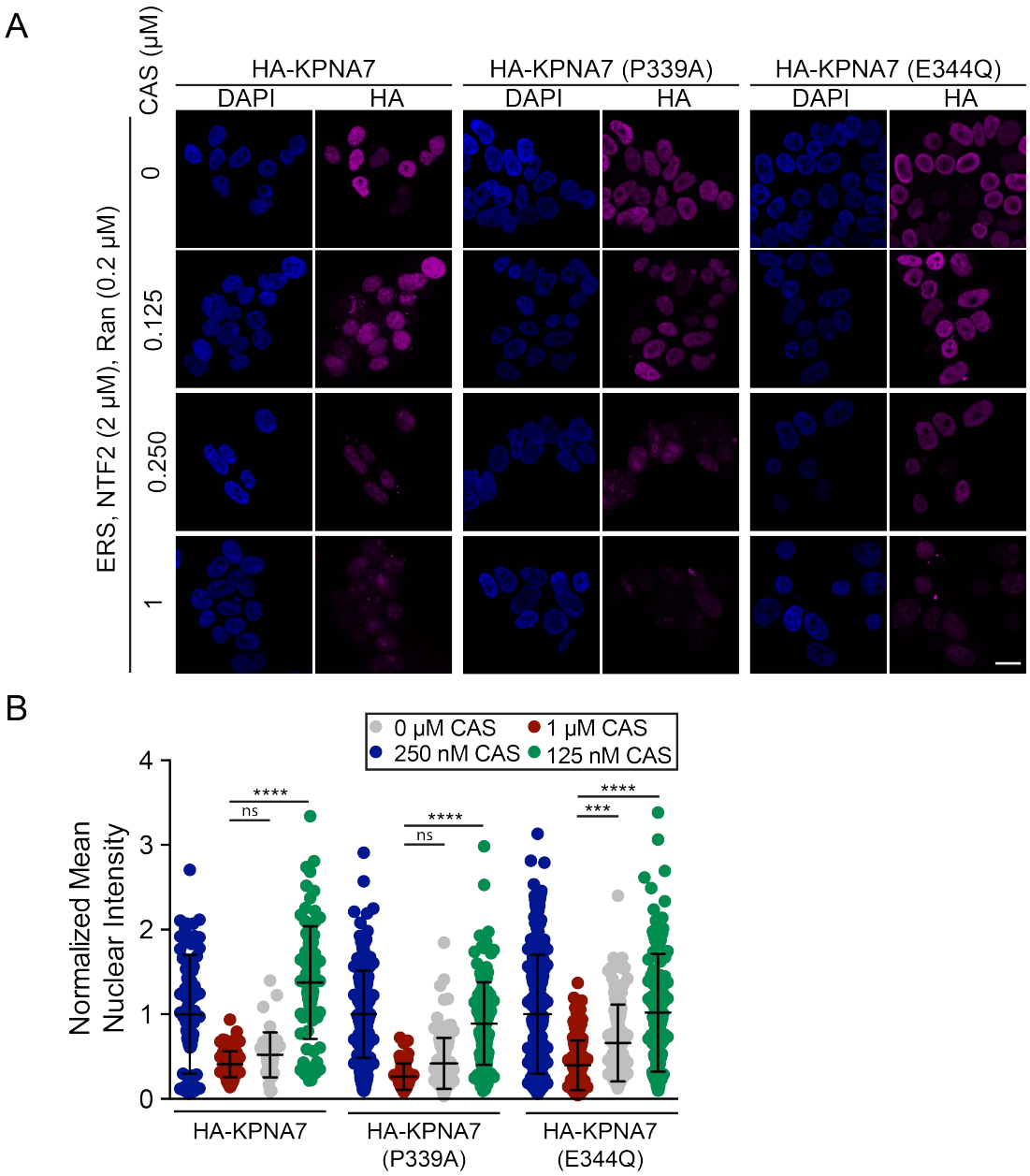


Figure S3.1: CAS-mediate export assays of KPNA7 and disease associated mutants. **(A)** Nuclear export of KPNA7 and KPNA7 mutants assayed in digitonin-permeabilized cells. HEK293T stably expressing HA-KPNA7 constructs were permeabilized with 0.005% digitonin, and incubated with CAS (concentrations indicated), Ran (0.2 μM), NTF2 (2 μM), and an energy regenerating system at 30°C. Post-export reactions were fixed and processed for IF microscopy for HA, and images were captured by confocal microscopy. (**B, C**) Mean nuclear intensity for each construct were determined using ImageJ and plotted. (**** $p<0.0001$, ***$p<0.001$; Scale: 10 μm)

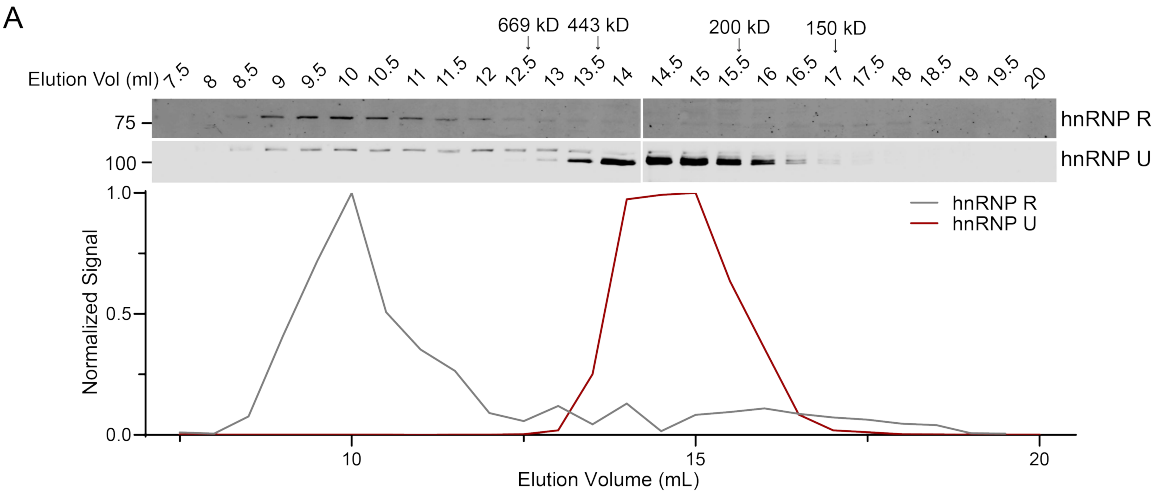


Figure S3.2: Size exclusion chromatography analysis of hnRNP proteins.
(A) HEK293T cell lysate was analyzed by size exclusion chromatography with a Superose6 column. Fractions were analyzed by SDS-PAGE and western blot analysis. Bands corresponding to hnRNP R and hnRNP U were quantified and plotted. Elution volume of standards of the indicated size are indicated above the blots.

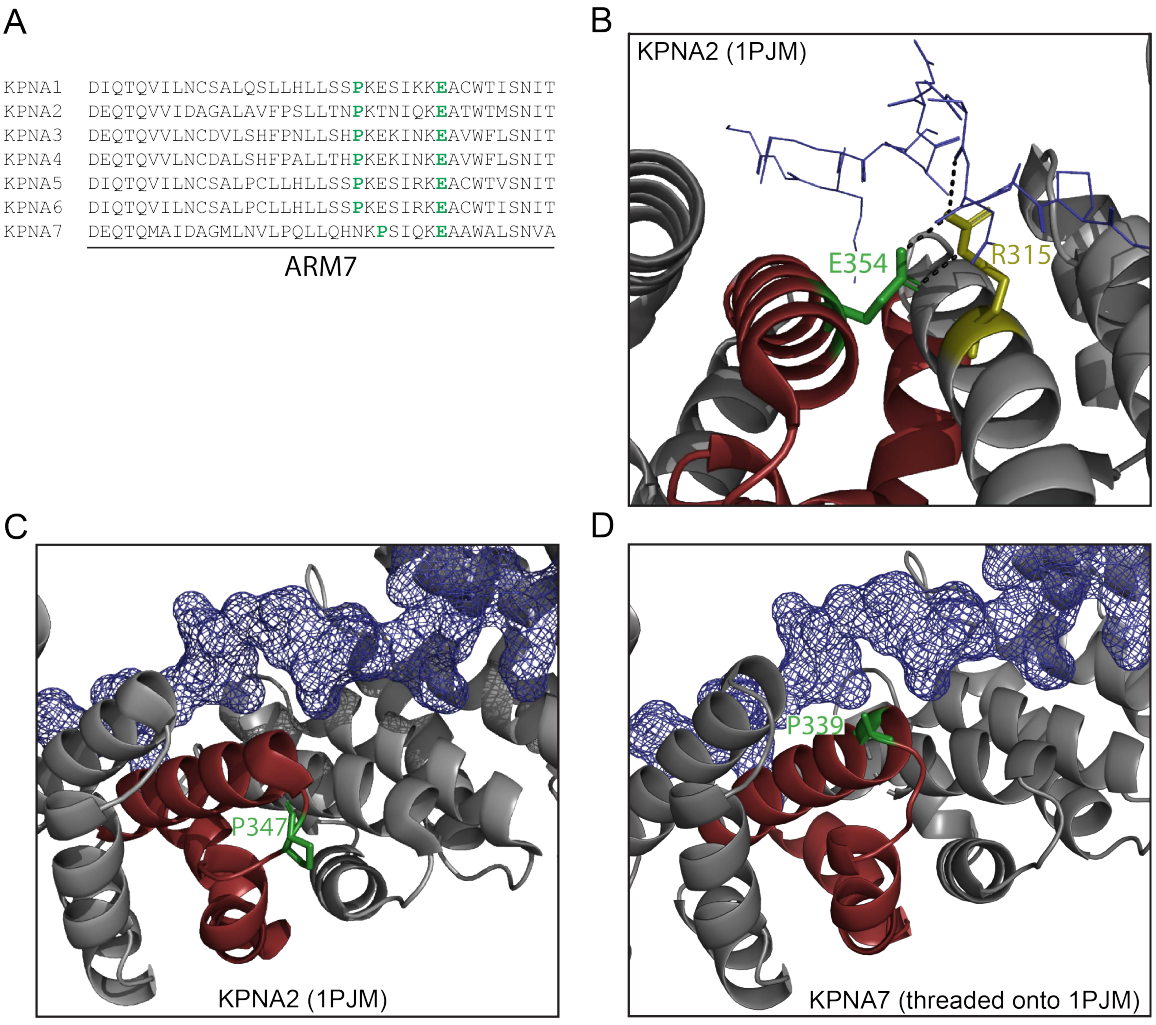


Figure S3.3: Locations of conserved residues in ARM7 of Imp-α.
(A) Multiple sequence alignment of ARM 7 of the seven human Imp-α isoforms. Highlighted in green are the 2 residues mutated in KPNA7 in epilepsy **(B)** Zoom in on E354 (green) in ARM 7 (red) of KPNA2 which interacts with R315 (yellow). Hydrogen bonds are shown in black dotted line. The bipartite Rb NLS is shown in blue. (PDB: 1PJM). **(C)** Highlight of the position of P347 (green) in KPNA2 between helices 2 and 3 of ARM 7 (red) (PDB: 1PJM). **(D)** Highlight of P339 in KPNA7 (primary sequence threaded onto the 1PJM KPNA2 structure with SWISS-MODEL). All images rendered in Pymol.

Table S3.1: KPNA7 interacting proteins identified by SILAC-MS with over 10 identified peptides and abundance ratios greater than 2.75

Description	Gene Symbol	# PSMs	Abundance Ratio: (F1,Heavy) /(F1,Light)
Isoform 2 of Heterogeneous nuclear ribonucleoprotein R OS=Homo sapiens OX=9606 GN=HNRNPR	HNRNPR	92	100
Isoform 2 of Serine/threonine-protein phosphatase PP1-alpha catalytic subunit OS=Homo sapiens OX=9606 GN=PPP1CA	PPP1CA	11	23.567
Pleiotrophin OS=Homo sapiens OX=9606 GN=PTN PE=1 SV=1	PTN	12	4.607
Histone deacetylase complex subunit SAP18 OS=Homo sapiens OX=9606 GN=SAP18 PE=1 SV=1	SAP18	36	4.017
RNA-binding protein with serine-rich domain 1 OS=Homo sapiens OX=9606 GN=RNPS1 PE=1 SV=1	RNPS1	38	3.99
Protein polybromo-1 OS=Homo sapiens OX=9606 GN=PBRM1 PE=1 SV=1	PBRM1	57	3.79
Cyclin-dependent kinase 11B OS=Homo sapiens OX=9606 GN=CDK11B PE=1 SV=4	CDK11B	24	3.531
Isoform SM-B1 of Small nuclear ribonucleoprotein-associated proteins B and B' OS=Homo sapiens OX=9606 GN=SNRPB	SNRPB	21	3.53
A-kinase anchor protein 17A OS=Homo sapiens OX=9606 GN=AKAP17A PE=1 SV=2	AKAP17A	23	3.384
Bromodomain-containing protein 9 OS=Homo sapiens OX=9606 GN=BRD9 PE=1 SV=2	BRD9	13	3.358
Splicing factor, arginine/serine-rich 19 OS=Homo sapiens OX=9606 GN=SCAF1 PE=1 SV=3	SCAF1	15	3.338
RNA-binding motif protein, X chromosome OS=Homo sapiens OX=9606 GN=RBMX PE=1 SV=3	RBMX	87	3.189
Bromodomain-containing protein 7 OS=Homo sapiens OX=9606 GN=BRD7 PE=1 SV=1	BRD7	12	3.183
Heterogeneous nuclear ribonucleoprotein U OS=Homo sapiens OX=9606 GN=HNRNPU PE=1 SV=6	HNRNPU	442	3.159
Pre-mRNA-splicing factor 38B OS=Homo sapiens OX=9606 GN=PRPF38B PE=1 SV=1	PRPF38B	15	3.126
U2 snRNP-associated SURP motif-containing protein OS=Homo sapiens OX=9606 GN=U2SURP PE=1 SV=2	U2SURP	67	3.124
Isoform 2 of Integrator complex subunit 3 OS=Homo sapiens OX=9606 GN=INTS3	INTS3	11	3.068
Zinc finger protein 768 OS=Homo sapiens OX=9606 GN=ZNF768 PE=1 SV=2	ZNF768	13	3.062
Peptidyl-prolyl cis-trans isomerase G OS=Homo sapiens OX=9606 GN=PPIG PE=1 SV=2	PPIG	24	3.061
40S ribosomal protein S11 OS=Homo sapiens OX=9606 GN=RPS11 PE=1 SV=3	RPS11	11	3.035
FACT complex subunit SSRP1 OS=Homo sapiens OX=9606 GN=SSRP1 PE=1 SV=1	SSRP1	10	3.014
Pinin OS=Homo sapiens OX=9606 GN=PNN PE=1 SV=5	PNN	57	2.934
E3 ubiquitin-protein ligase Hakai OS=Homo sapiens OX=9606 GN=CBLL1 PE=1 SV=1	CBLL1	10	2.916
Pre-mRNA-processing factor 40 homolog A OS=Homo sapiens OX=9606 GN=PRPF40A PE=1 SV=2	PRPF40A	86	2.912
Serine/arginine-rich splicing factor 9 OS=Homo sapiens OX=9606 GN=SRSF9 PE=1 SV=1	SRSF9	31	2.902
Cleavage and polyadenylation specificity factor subunit 2 OS=Homo sapiens OX=9606 GN=CPSF2 PE=1 SV=2	CPSF2	26	2.9
Protein virilizer homolog OS=Homo sapiens OX=9606 GN=VIRMA PE=1 SV=2	KIAA1429	32	2.887
Splicing factor U2AF 35 kDa subunit OS=Homo sapiens OX=9606 GN=U2AF1 PE=1 SV=3	U2AF1	95	2.883
Isoform 4 of Apoptotic chromatin condensation inducer in the nucleus OS=Homo sapiens OX=9606 GN=ACIN1	ACIN1	126	2.881
Nucleolar protein 4 OS=Homo sapiens OX=9606 GN=NOL4 PE=1 SV=2	NOL4	12	2.859
General transcription factor IIF subunit 2 OS=Homo sapiens OX=9606 GN=GTF2F2 PE=1 SV=2	GTF2F2	27	2.849

U4/U6.U5 tri-snRNP-associated protein 2 OS=Homo sapiens OX=9606 GN=USP39 PE=1 SV=2	USP39	20	2.816
Transcription elongation factor SPT5 OS=Homo sapiens OX=9606 GN=SUPT5H PE=1 SV=1	SUPT5H	36	2.797
Pre-mRNA-splicing factor 38A OS=Homo sapiens OX=9606 GN=PRPF38A PE=1 SV=1	PRPF38A	11	2.795
RNA-binding protein 39 OS=Homo sapiens OX=9606 GN=RBM39 PE=1 SV=2	RBM39	59	2.779
Isoform 2 of Heterogeneous nuclear ribonucleoprotein U OS=Homo sapiens OX=9606 GN=HNRNPU	HNRNPU	393	2.774
Protein DEK OS=Homo sapiens OX=9606 GN=DEK PE=1 SV=1	DEK	60	2.774
Zinc finger RNA-binding protein OS=Homo sapiens OX=9606 GN=ZFR PE=1 SV=2	ZFR	10	2.76
60S ribosomal protein L24 OS=Homo sapiens OX=9606 GN=RPL24 PE=1 SV=1	RPL24	10	2.754
SAP30-binding protein (Fragment) OS=Homo sapiens OX=9606 GN=SAP30BP PE=1 SV=1	SAP30BP	14	2.75

Chapter IV

Towards improving proximity labeling by the biotin ligase BirA

Luke T. Oostdyk, Leonard Shank, Kasey Jividen, Natalia Dworak, Nicholas E. Sherman, Bryce M. Paschal

Published: Methods, 15 March 2019, Volume 157, Pages 66-79

I. Abstract

The discovery and validation of protein-protein interactions provides a knowledge base that is critical for defining protein networks and how they underpin the biology of the cell. Identification of protein interactions that are highly transient, or sensitive to biochemical disruption, can be very difficult. This challenge has been met by proximity labeling methods which generate reactive species that chemically modify neighboring proteins. The most widely used proximity labeling method is BioID, which features a mutant biotin ligase BirA(Arg118Gly), termed BirA*, fused to a protein of interest. In this chapter, we explore how amino acid substitutions at Arg118 affect the biochemical properties of BirA. We found that relative to BirA*, the Arg118Lys substitution increased biotin affinity and release of reactive biotinyl-5-AMP. BioID using a BirA(Arg118Lys)-lamin A fusion enabled identification of PCNA as a lamina-proximal protein in HEK293T cells, a finding that was validated by immunofluorescence microscopy. Our data expand on the concept that proximity labeling by BirA fused to proteins of interest can be modulated by amino acid substitutions that affect biotin affinity and the release of biotinyl-5’-AMP.

II. Introduction

A. *Enzyme-based methods for identifying protein-protein interactions in cells*

The identification of protein-protein interactions (PPIs) in eukaryotic cells is critical for understanding basic protein function as well as the organization and regulation of cellular pathways. A major challenge in this regard is that many protein-protein interactions are transient and/or characterized as low affinity. Widely-used biochemical techniques such as immunoprecipitation and *in vitro* binding assays, which can provide protein-protein interaction data, rely on interactions that are sufficiently stable or can be reconstituted outside of the environment of the cell. In recent years, methods that provide a snapshot of protein-protein interactions in the cell have been developed. Proximity-dependent labeling techniques utilize enzymes fused to a protein of interest (POI) that can transfer a molecular tag to proteins situated within a suitable labeling radius in the cell (Chen and Perrimon, 2017; Li et al., 2017; Rees et al., 2015). The molecular tag facilitates enrichment and identification of the labeled species, the latter usually by mass spectrometry (MS). Proximity-dependent labeling techniques can therefore overcome some of the difficulties with studying transient and/or weak protein-protein interactions in the cell, as well as the challenges associated with dissecting macromolecular complexes since the labeling occurs when cellular architecture is intact (**Fig. 4.1A**).

Most proximity-dependent labeling approaches utilize biotin, or biotin conjugates, as the tag, and streptavidin for detection and enrichment of labeled species. To date, two types of enzymes have been used, peroxidases and biotin-ligases. The peroxidases employed are horseradish peroxidase (HRP) and ascorbate peroxidase (APEX), while the biotin-protein ligase used for proximity labeling is the BirA enzyme from *E. coli*. HRP and APEX can, in the presence of hydrogen peroxide (H_2O_2), convert a phenolic substrate into a highly reactive radical. Phenolic substrates such as biotin-arylazides (Honke and Kotani, 2012), biotin-tyramides (Li et al., 2014), and biotin-phenols (Hung et al., 2014) can be used for proteomic applications, and 3,3'-diaminobenzidine can be used to generate an electron-dense product for visualization by electron microscopy (Hopkins et al., 2000). Radicals released by both HRP and APEX are short lived (<1 ms), but react with electron-rich amino acids including Tyr, Trp, His and Cys with a predicted labeling

radius of 20 nm (Rhee et al., 2013). HRP-based labeling has been used to identify protein-protein interactions on the cell surface (Jiang et al., 2012), but HRP is inactive in the reducing environment of the cytosol. The development of APEX circumvented this problem. The engineered form of APEX is functional over a broad pH range, and shows enzymatic activity in several cellular compartments of mammalian cells including the mitochondrial matrix, endoplasmic reticulum, cytoplasm, and nucleus (Martell et al., 2012).

The most widely used proximity-dependent labeling technique utilizes a mutant form of the biotin-protein ligase BirA. In *E. coli* and other bacteria, BirA regulates the enzyme activity of acetyl-CoA carboxylase through biotinylation of the carboxyl carrier protein (BCCP) subunit (Cronan, 1989). BirA also helps coordinate biotin synthesis by regulating transcription of the biotin biosynthetic operon (Cronan, 1989; Nenortas and Beckett, 1996; Otsuka and Abelson, 1978). The mechanism of BirA biotinylation has been studied in detail. Briefly, BirA binds biotin and ATP and generates the intermediate, biotinoyl-5'-AMP (bioAMP) (McAllister and Coon, 1966). At low biotin concentrations, bioAMP is transferred to the amine group of a lysine in BCCP. Under conditions of high biotin concentration, where all BCCP subunits are biotinylated, bioAMP remains bound to BirA; *in vitro* the BirA-bioAMP complex has a half-life of ~30 min (Xu and Beckett, 1994). In bacteria, the BirA-bioAMP complex binds to a 40-base pair sequence in the biotin biosynthetic operon, resulting in the transcriptional repression of genes encoding enzymes required for biotin biosynthesis (Otsuka and Abelson, 1978).

The premise of identifying protein interactions through proximity labeling by BirA is based on structure-function work from Cronan and others. Amino acid substitutions in the biotin-binding site of BirA, notably Arg118Gly, were shown to dramatically increase the bioAMP dissociation rate (Kwon and Beckett, 2000; Weaver et al., 2001). Cronan suggested bioAMP release could provide the basis for promiscuous chemical biotinylation of lysine residues within proteins in the solvent (Choi-Rhee et al., 2004; Cronan, 2005). Proof-of-concept came with the observations that recombinant BirA(Arg118Gly) could mediate biotinylation of purified proteins *in vitro*, and that multiple proteins in *E. coli* underwent promiscuous biotinylation in the presence of this

mutant enzyme (Choi-Rhee et al., 2004). The BirA(Arg118Gly) mutant was designated BirA*. The Roux laboratory published an elegant implementation of the method in mammalian cells where they showed that BirA* fused to Lamin A mediates proximal labeling of proteins within the nuclear lamina. These authors coined the term BioID to describe the technology (Roux et al., 2012). When tethered to a nuclear pore complex protein, BirA* was shown to have an effective labeling radius of 10 nm (Kim et al., 2014). This is an approximation based on high resolution mapping of nucleoporin positions within the nuclear pore complex. The effective labeling radius is presumably generated by a high local concentration of lysine-reactive bioAMP, but the labeling is likely influenced by multiple factors including how the POI is anchored within a macromolecular structure, the flexibility afforded by BirA* tethering, and the density of solvent-exposed lysine residues in proximity. BirA* fused to a POI can generate compartment-specific targeting and biotinylation of proteins, which could include proteins that bind directly or indirectly to the POI (**Fig. 4.1A**). Enrichment of the biotinylated species and MS analysis provides an inventory of proteins labeled by bioAMP released from BirA* in the compartment or structure of interest.

To date, BioID has been used successfully to study a variety of macromolecular structures. Examples include BirA* targeting to the nuclear lamina (Fu et al., 2015; Mehus et al., 2016; Roux et al., 2012), the nuclear pore complex (Kim et al., 2014), cell junctions (Fredriksson et al., 2015; Van Itallie et al., 2013, 2014), and centrosomes (Comartin et al., 2013; Firat-Karalar and Stearns, 2015; Firat-Karalar et al., 2014; Gupta et al., 2015). BirA* has also been targeted to membrane-bound compartments such as the mitochondrial matrix (Cole et al., 2015), and it has also been used in contexts where the POI fusion is not tethered to a cellular structure (Cheng et al., 2015; Couzens et al., 2013; Sun et al., 2015; Zhou et al., 2015).

B. *Dependence on and limitations of the biotin-avidin interaction*

The selective, high affinity interaction between biotin and avidin (and its derivatives; $K_d = 10^{-14}$ M; (Green, 1975)) is a critical feature of BioID. Enrichment for biotinylated species is specific and tolerates the stringent wash conditions necessary to dissociate unlabeled proteins and contaminating macromolecular structures including cellular

membranes. The wash stringency is important because in most published datasets, identification of proteins enriched on streptavidin beads by MS does not reveal a biotin mass addition (+226.08; biotinyl-lysine). In other words, MS typically scores peptides (masses) derived from the protein enriched without identifying the biotin-containing site responsible for enrichment. The fact that biotinylated peptides are not usually detected might reflect (i) the inefficient release of biotinylated peptides from streptavidin; (ii) biotin modification of lysine residues interfering with trypsin recognition of the cleavage site; and (iii) a low stoichiometry of biotin labeling. Thus, the proximity assessment of proteins is usually inferred from counting the number of peptides derived from a protein enriched on streptavidin beads. It is worth noting that one study succeeded in direct identification of biotinylated peptides using a strategy termed Biotin Site Identification Technology (BioSITe) (Kim et al., 2018a). In this method biotinylated proteins are digested with trypsin prior to enrichment with an immobilized anti-biotin antibody, which has a lower affinity for biotin than streptavidin. This approach helps eliminate non-biotinylated peptides as contaminates from the pipeline, it allows for more efficient elution of biotinylated peptides for MS analysis, and it enables MS detection of the biotin modification on peptides. A similar strategy involves using Strept-Tactin for enrichment, elution, and MS identification of biotinylated peptides (Opitz et al., 2017).

C. *Refining BirA-based proximity labeling*

While BioID with BirA* has been highly successful, there have been efforts to improve the technology. A second-generation technique termed BioID2 was developed by Roux and colleagues based on a lower molecular weight biotin ligase from *Aquifex aeolicus* (Kim et al., 2016). The smaller ligase (26 vs. 35 kD) was shown to improve targeting of protein fusions as well as mediate labeling with a lower biotin concentration in the culture media. Very recently, the Ting laboratory has used yeast display-based directed evolution to generate two BirA mutants, TurboID and miniTurbo, which rapidly release bioAMP and can be used for proximity dependent biotinylation in as little as 10 minutes in mammalian cells (Branon et al., 2018).

We took a highly focused approach that involved generating amino acid substitutions in Arg118 in the biotin-binding site of BirA. We focused on Arg118 in BirA, which in the

crystal structure (PDB: 1HXD) is positioned to make polar contacts with the carboxyl group of biotin (yellow) and with Arg121 in the biotin binding loop (**Fig. 4.1B**; (Weaver et al., 2001)). The Gly substitution for Arg118 in BirA* likely alters the loop position since this reduces biotin affinity by 200-fold (Kwon and Beckett, 2000). We reasoned that other amino acid substitutions at this position might alter bioAMP dissociation (positively or negatively) and impact biotinylation. We found that BirA(Arg118Lys) can mediate proximity labeling in cells under low biotin concentrations, and in vitro, can be inactivated by auto-modification at a high biotin concentration.

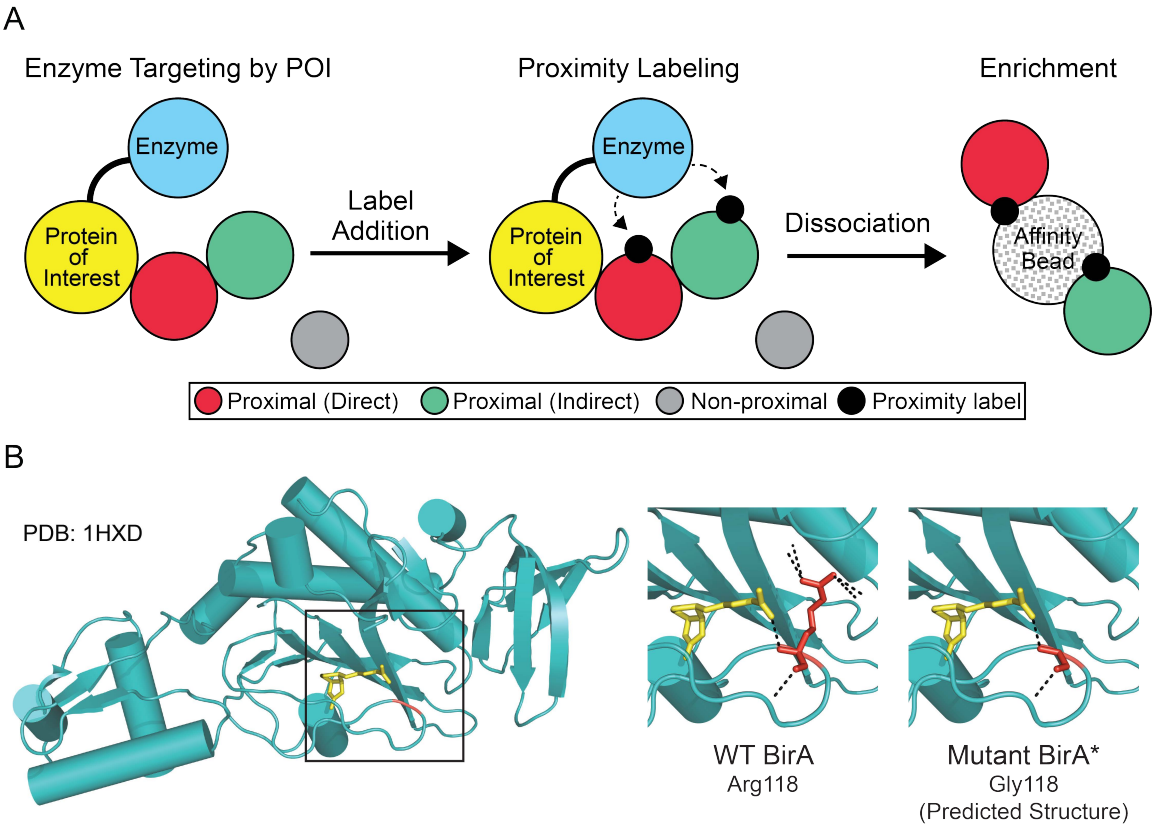


Figure 4.1: Proximity ligation protein labeling.
(A) Overview of enzyme-based proximity labeling. An enzyme capable of generating a reactive intermediate is expressed as a fusion with a protein of interest (yellow) that results in appropriate subcellular targeting. Label addition results in proximal labeling of proteins bound directly (red) and indirectly (green) to the protein of interest. Proteins are subsequently solubilized, and the proximity-modified proteins are enriched on beads by virtue of the label. The proteins labeled in this manner are identified by mass spectrometry. **(B)** Crystal structure of BirA (PDB: 1HXD) with biotin (yellow) bound to the active site. Residue 118 is highlighted in red with the biotin binding site boxed in black. The panels to the right show the biotin binding site of BirA and the hydrogen bonds (black dotted lines) associated with Arg118 and Gly118. Modeling performed in PyMol.

III. Materials and Methods

A. Expression, purification, and assay of BirA

BirA proteins were expressed and purified as Maltose Binding Protein (MBP) fusions. The MBP tag was selected as the fusion partner because it affords high levels of expression and solubility in *E. coli*, and it contains an abundance of lysine residues (33) that could potentially serve as promiscuous biotinylation sites. The BirA coding region was cloned into the pMal-p2x vector (NEB) and site-directed mutagenesis (QuickChange II; Agilent Technologies #200523) was performed to generate plasmids encoding 19 different amino acids at position 118.

BirA proteins were expressed in BL21 *E. coli* in LB broth and purified with amylose resin (NEB; E8021S). In brief, MBP-BirA plasmids were transformed into chemically competent BL21 *E. coli* (NEB; C2530), plated on LB-ampicillin plates, and incubated at 37°C for 16 hours. Starter cultures were made by scraping ~10 colonies into LB medium (25 ml) containing ampicillin (100 µg/ml) and shaken at 37 °C for 2 hours. The starter culture was then added to LB amp (1 L) with 0.2% glucose and shaken at 37°C to an OD_{600} ~0.6. Cultures were cooled on ice for 10 minutes, induced with IPTG (1 mM), and shaken at 18°C for 18 hrs. Prior to harvest, the cultures were supplemented with PMSF (1 mM) and shaken briefly. Cultures were collected by centrifugation (4300g for 15 minutes at 4°C). Pellets were resuspended in 12.5 ml of amylose column buffer (20 mM Tris pH 8, 200 mM NaCl, 1 mM EDTA, 1 mM DTT, and 1 µg/ml each of leupeptin, pepstatin, and aprotinin). Cells were lysed by passage through a French press (3X). After clarification by centrifugation (30,000g for 30 minutes at 4°C), the soluble extract (~15 ml) was combined with amylose resin (1 ml packed beads) in a chromatography column (27 ml; Kimble, 420400-1515). The column was rotated end-over-end (4°C, 2 hrs), the resin allowed to settle, and flow-through was collected. Beads were then washed with amylose column buffer (50 ml), and MBP fusions subsequently eluted with 10 mM maltose in amylose column buffer. Peak protein fractions were determined by A_{280} measurement, and fractions were pooled and dialyzed into PBS containing DTT (1 mM). The dialyzed MBP-BirA proteins were then supplemented with protease inhibitors (1 µg/ml each of leupeptin, pepstatin, and aprotinin, 1 mM PMSF). The purified proteins (0.4-2 mg/ml, depending on the mutant) were dispensed into small aliquots, snap frozen

in liquid nitrogen, and stored at -80 °C. Most experiments were performed using single-use aliquots, though there was no apparent loss of activity when BirA proteins were stored on ice for several days.

It has been previously shown that BirA* undergoes auto-biotinylation in *E. coli* (Choi-Rhee et al., 2004). This provides a simple readout for activity of BirA mutants. To determine the auto-modification state, each BirA protein (1 µg) was resolved by SDS-PAGE, transferred to a nitrocellulose membrane, and detected with fluorescent-neutravidin (fl-neutra; ThermoFisher, 22853) by LI-COR Odyssey Clx infrared imager. A duplicate gel was stained with Coomassie Blue to ensure equal loading of the BirA proteins (**Fig. 4.2A**). From this analysis, we learned that different amino acid substitutions for Arg at position 118 can modestly increase or decrease biotin conjugation to MBP-BirA when expressed in *E. coli*.

The biotinylation signal from the MBP-BirA fusion protein could reflect BirA auto-modification, or BirA labeling of MBP. BirA and MBP contain 18 and 33 Lys residues, respectively, any of which could be reactive with bioAMP. To distinguish between BirA auto-modification and MBP biotinylation, we subjected the recombinant fusion proteins to proteolytic cleavage using the Factor Xa cleavage site positioned between MBP and BirA. Following cleavage, the biotinylation state was assessed by probing with fl-neutra (**Fig. 4.2B**).

For cleavage reactions, MBP-BirA proteins (5 µg) were each dialyzed into Factor Xa cleavage buffer (20 mM Tris-HCl pH 8, 100 mM NaCl, 2 mM $CaCl_2$). The dialyzed proteins were then incubated with Factor Xa (0.5 µg; NEB; P8010S) at room temperature for 16 hours on a benchtop shaker. Following this reaction, cleaved (2.5 µg) and un-cleaved (2.5 µg) mutant enzyme was separated on an 8% SDS-PAGE gel, transferred to nitrocellulose membrane and detected with fl-neutra and anti-MBP antibody (NEB; E8032S) with anti-mouse AlexaFluor680 secondary antibody (ThermoFisher, A28183) (**Fig. 4.2C**). We found differences in the relative levels of labeling of MBP and BirA that were dependent on the amino acid substitution. BirA(WT) displayed a very low level of biotinylation that was restricted to MBP. BirA* (Arg118Gly substitution) showed labeling on both MBP and BirA, but was biased towards modification of BirA itself. We

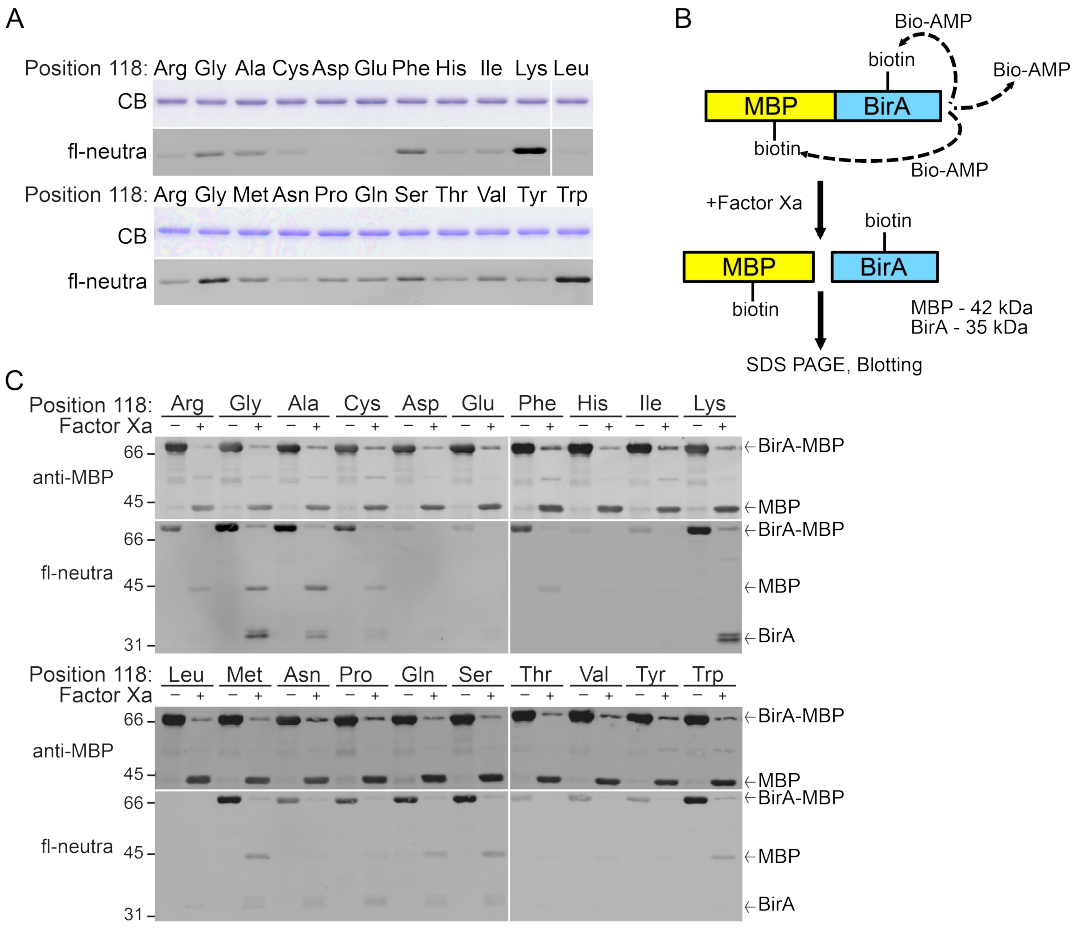


Figure 4.2: Evaluation of auto-biotinylation by mutants of BirA at position 118.
(A) Purification and biotinylation state (fl-neutra detection) of wild-type and mutant BirA proteins expressed as MBP fusions in *E. coli*. Each mutant protein contains a single amino acid substitution at position 118, as indicated. Equal amounts (1 µg) of each MBP-BirA mutant were analyzed by Coomassie blue staining and by blotting. **(B)** Scheme for Factor Xa mediated cleavage and analysis of auto-biotinylation. **(C)** Biotin detection within the MBP and BirA catalytic fragments of each BirA mutant (2.5 µg) before and after cleavage with Factor Xa. Blots were also probed with anti-MBP antibody. Bands corresponding to MBP-BirA, MBP, and BirA are indicated to the right of each panel.

observed that Arg118Gly and Arg118Lys showed the strongest level of automodification (**Fig. 4.2C**), and in both mutants BirA appeared as a doublet with slightly higher labeling of the faster migrating species. We also noted the Arg118Trp mutant biotinylated MBP but this occurred without obvious automodification. Based on these data, we set out to study the Arg118Lys and Arg118Trp mutants. We subsequently determined from transfection experiments that LaminA-BirA (Arg118Trp) mutant has very low biotinylation activity in cells even with biotin supplementation (unpublished observations). These data led us to a comparative analysis of BirA(Arg118Lys), BirA(Arg118Gly) and BirA(WT).

A. Bead-based biotinylation assay with BirA

We designed a simple bead-based assay with the goal of testing whether BirA tethered to a surface could mediate promiscuous biotinylation of a substrate in trans (**Fig 4.3A**). MBP-BirA fusion proteins (WT, Arg118Gly, and Arg118Lys; 5 µg each) and MBP (10 µg, corresponding to a 4-fold molar excess over MBP-BirA) were immobilized on amylose resin (10 µl packed beds) by rotation (4 hrs, 4°C). We also immobilized MBP alone (10 µg) in the absence of BirA, which serves as a blotting control. Beads were then washed four times with *in vitro* biotinylation buffer (IVBB) (50 mM Tris pH8, 5 mM $MgCl_2$, 1 mM DTT, 1 µg/ml each aprotinin, leupeptin, and pepstatin) and resuspended with IVBB (20 µl) and placed on ice. ATP (3 mM final) and biotin (10 µM) were added, and the reaction was supplemented with IVBB (50 µl total volume). Samples were then incubated at 37°C for 30 minutes. The reactions were chilled briefly on ice, and then centrifuged to pellet the beads. The supernatant was aspirated, beads were washed four times with IVBB (300 µl), and 1X SDS-PAGE sample buffer (50 µl) was added to each tube. The sample (20%) was analyzed by SDS-PAGE and immunoblotting with anti-MBP antibody to confirm equivalent loading, and fl-neutra was used to detect biotin labeling (**Fig. 4.3B**). MBP and fl-neutra signals were quantified using Image Studio (LI-COR) or ImageJ software. Biotinylation was plotted by normalizing to the level of each protein (MBP signal) for MBP-BirA (cis-biotinylation) and MBP (trans-biotinylation) (**Fig. 4.3C**). The level of trans-biotinylation by the BirA(Arg118Lys) mutant was higher than that of BirA*. Surprisingly, BirA(WT) was also capable of labeling MBP in trans. This suggests that even the WT enzyme is capable of releasing sufficient bioAMP for

biotinylation of a proximal protein. It should be noted that non-enzymatic (BirA-independent) biotinylation of MBP is not observed in this assay.

B. BirA activity measured using biotin acceptor peptides

To further characterize the biotin ligase activity of the BirA(Arg118Lys) mutant, we examined its ability to label a biotin acceptor peptide (BAP). We used a 23 amino acid BAP isolated by phage display (Schatz, 1993) and expressed it as a GST fusion protein. The single lysine embedded in the BAP sequence (**Fig. 4.4A,** bold underlined) serves as the biotin acceptor in a reaction that requires addition of ATP and biotin. For controls, the acceptor lysine was changed to an alanine (K>A) or arginine (K>R), the latter to retain the positive charge (**Fig. 4.4A**). For the TLC experiments described below, synthetic peptides corresponding to the WT, Lys to Ala, and Lys to Arg sequences (**Fig. 4.4A**) were purchased (GenScript). Peptides were resuspended in water (1.5 mM) and frozen (-80°C).

BAP amino acid sequences (WT, Lys to Ala) were reverse translated and codon optimized for expression in *E. coli*. Single strand oligonucleotides were purchased with overhangs to permit direct cloning into the pGEX-4T1 plasmid after restriction digest with EcoRI and XhoI. Thus, the sense strand contained a 5' overhang (5'-AATT), and the antisense strand contained a 5' overhang (5'-TCGA). The pGEX-4T1 plasmids were transformed into chemically competent BL21 *E. coli*, plated on LB-ampicillin plates, and grown at 37°C for 16 hours. Because *E. coli* contain endogenous BirA that biotinylates the GST-BAP substrate during growth, cultures are maintained in M9 minimal medium lacking biotin. M9 minimal medium is prepared as follows: (i) A 5X M9 minimal salts solution is prepared by dissolving $NaHPO_4$ (30 g), KH_2PO_4 (15 g), NaCl (2.5 g) and NH_4Cl (5 g) in H_2O (1 L) followed by sterile filtration. (ii) Separately, glucose (4 g) is added to H_2O (750 ml) and autoclaved. (iii) $MgSO_4$ (1 mM final), $CaCl_2$ (1 mM final) and 5X M9 salts (200 ml) are added to the glucose solution with sterile water for a final volume of 1 L.

Starter cultures were made by scraping ~10 colonies into M9 minimal medium (25 ml) containing ampicillin (100 μg/ml). The starter culture was shaken at 37 °C for 2 hours, and then added to the growth culture in the same medium (1 L) and shaken at 37°C until

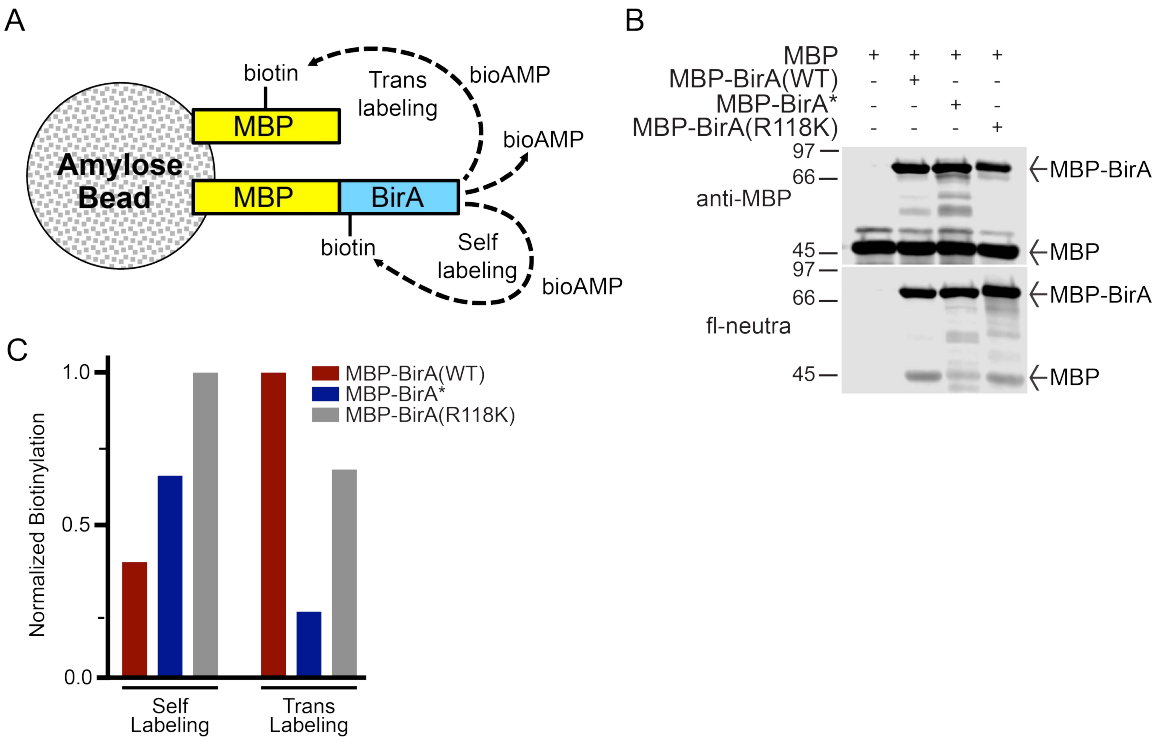


Figure 4.3: Bead-based analysis of proximal biotinylation by BirA mutants.
(A) Scheme for proximal (self and trans) biotinylation by BirA immobilized on an amylose bead surface. **(B)** Biotinylation reactions were performed with MBP-BirA mutants (5 µg) and free MPB (10 µg) on amylose beads in the presence of biotin (50 µM). Samples were analyzed by fl-neutra. Normalized biotinylation of self and trans labeling was quantified and plotted in **(C)**.

the OD_{600} reaches 0.6. The growth culture was induced with IPTG (1 mM) and EtOH (2%) and shaken for 3 hours at 37°C. Just prior to harvest, PMSF (1 mM) was added to the culture. The culture was collected by centrifugation (4300g for 15 min at 4°C). The pellets were resuspended in PBS with DTT (1 mM), with leupeptin, pepstatin, and aprotinin (1 µg/ml each) and lysed by passing through a French press 3 times. The lysate was clarified by centrifugation (30,000g 30 min at 4°C). Triton X-100 (using a 20% stock) was added to the soluble fraction to a final concentration of 1%, and the sample (12.5 ml) was combined with glutathione agarose (1 ml packed beads; Sigma; G4510). The sample was mixed by end-over-end rotation in a chromatography column (27 ml) at 4°C (2 hrs). The beads were allowed to settle, the column flow-through was collected, and beads were washed sequentially with the following buffers: (1) PBS with 1% Triton X-100 (25 ml), (2) PBS with 500 mM NaCl (25 ml) and (3) PBS (25 ml). GST-BAP is eluted in glutathione (10 mM), 50 mM Tris, pH 8 (50 mM), DTT (1 mM), with leupeptin, pepstatin, and aprotinin (1 µg/ml each). The fractions were measured by A_{280}, and peak fractions were selected, pooled, and dialyzed into PBS containing DTT (1 mM). Protease inhibitors (leupeptin, pepstatin, and aprotinin; 1 µg/ml each) were added after dialysis and proteins were aliquoted and stored at – 80 °C.

C. BirA substrate recognition and promiscuity analyzed using BAPs

The purified GST-BAP fusions are used to examine the biotin acceptor site-specific and potentially promiscuous biotinylation by BirA proteins. In this experiment, the GST-BAP mutant (K>A) serves as a substrate for promiscuous biotinylation by BirA because the single lysine in the 23 amino acid BAP is changed to a non-biotinylated amino acid, alanine. Thus, any biotinylation observed with GST-BAP mutant (K>A), occurs as a result of promiscuous labeling on one or more of the 21 lysine residues in GST.

For each reaction, an MBP-BirA fusion (1 µg) was combined with either GST-BAP (50 ng), GST-BAP(K>A) (1 µg), or GST-BAP(K>A) (10 µg). Samples were prepared in a total volume of 50 µL in IVBB with ATP (3 mM) and biotin (50 µM) on ice. Reactions were incubated at 37°C for 1 hour and stopped with the addition of 5X SDS-PAGE sample buffer (10 ul). Samples (20% of total reaction) were analyzed by SDS-PAGE and immunoblotting with anti-MBP and anti-GST (Santa Cruz; sc-138) antibodies along with

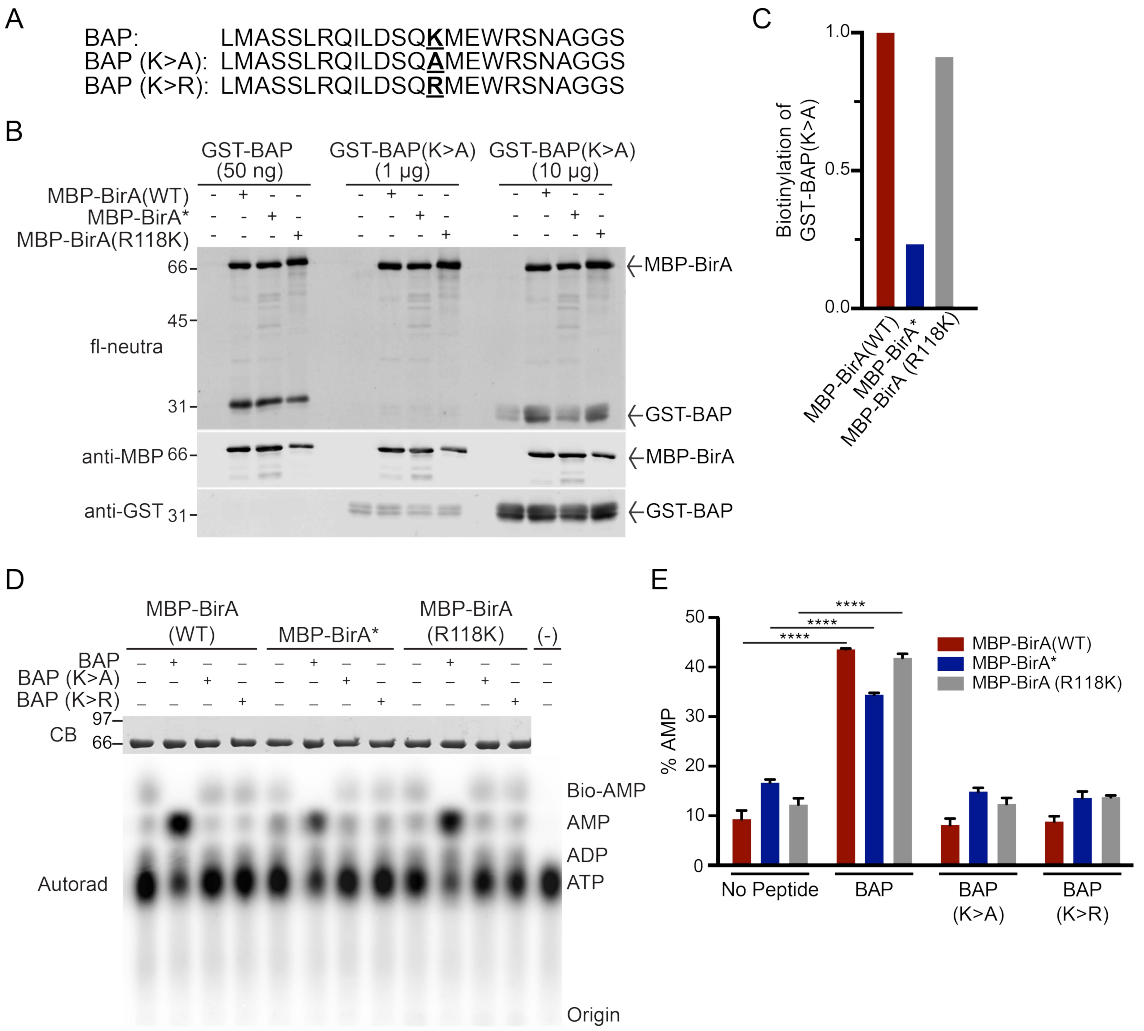


Figure 4.4: Utilization of Biotin Acceptor Peptide (BAPs) for analyzing BirA mutants. **(A)** BAP sequences showing the position of the acceptor lysine and amino acid substitutions (bold underline). **(B)** *In vitro* biotinylation of BAP substrates with purified BirA enzymes. Reactions contained MBP-BirA (1 µg) with the indicated amount of each BAP and biotin (50 µM). **(C)** Quantification of relative biotinylation of GST-BAP (K->A) in (B) indicates proximity labeling of a site outside of the BAP acceptor lysine. **(D)** TLC analysis of basal and BAP-induced Bio-AMP and AMP generation by BirA proteins. Assays were performed in triplicate and analyzed by autoradiography, one set of which is shown in this panel. **(E)** Quantification of AMP generation by BirA mutants. The % AMP is the percentage of signal in each lane corresponding to [α-^{32}P]-labeled AMP. (****p<0.0001)

fl-neutra to detect biotinylation (**Fig. 4.4B**). Quantification was performed using Image Studio or ImageJ software where fl-neutra signal from samples with 10 µg of GST-BAP(K>A) were quantified by subtracting signal from the negative control and normalizing to each BirA enzyme (**Fig. 4.4C**). The results from this experiment indicated that all three forms of BirA biotinylate the WT BAP substrate when it is used at a low concentration (50 ng). Virtually no labeling of the mutant BAP (K>A) was detected when it is used at 20-fold higher concentration than the WT BAP. However, the mutant BAP did show a low level of labeling when used at the highest concentration (10 ug) by BirA(WT) and BirA(R118K). These data indicated that for this particular substrate, the WT and Arg118Lys mutant BirA mediate a higher level of promiscuous labeling of a mutant BAP substrate than the Arg118Gly mutant, BirA*.

To better characterize biotinylation mediated by the Arg118Gly and Arg118Lys mutants, we utilized thin layer chromatography (TLC) to monitor the chemical products of the reaction (Chakravartty and Cronan, 2012). The ATP- and biotin-dependent generation of bio-AMP by BirA is followed by AMP release and biotin conjugation to BAP. The use of [α-^{32}P]-labeled ATP in biotinylation reactions with TLC enables the visualization of ATP, ADP, AMP, and bio-AMP by autoradiography (autorad) because of the different mobilities of these chemical species. In these reactions, synthetic peptides encoding WT and mutant BAPs are added and the relative level of AMP production are determined in order to assess how well different BirA enzymes respond to substrate.

Biotinylation reactions for analysis by TLC were performed in a modified IVBB containing Tris, pH 8 (5 mM), TCEP (0.5 mM), $MgCl_2$ (0.5 mM), KCl (10 mM), ATP (50 µM), and [α-^{32}P] ATP (82.5 nM; Perkin Elmer). Each reaction (15 ul total) contained MBP-BirA (2 µM), BAP peptide (50 µM), and biotin (50 µM). Reactions were assembled on ice and incubated at 37°C for 1 hour. In duplicate, 2 ul of each reaction was spotted with 1 cm spacing exactly 2 cm from the base of a microcrystalline cellulose matrix plate (20 cm x 20 cm; Analtech, 05011) and placed in a chamber with ~1 cm depth of an isobutryic acid:NH_4OH:water (66:1:33 by volume) solution. The buffer front was allowed to migrate ~15 cm (three quarters of the plate height) which typically takes between 1-2 hours. The plate was removed from the chamber and allowed to air dry. The

plate was then covered with clear plastic wrap and imaged by exposure to X-ray film and/or a Phosphor Imager (Molecular Dynamics). One-half of each reaction (7.5 µl) was also analyzed by SDS-PAGE and Coomassie blue staining to confirm equal loadings (**Fig. 4.4D**). Reactions were spotted on duplicate plates as technical replicates. Utilizing the plot lanes function in ImageJ, the signal from ATP, ADP, AMP, and bioAMP was quantified. The total signal in each lane was calculated and the percentage of signal corresponding to AMP was plotted (**Fig. 4.4E**). Percent AMP is a measure of turn-over of bio-AMP to biotin and AMP in the active site of BirA. It was observed that all three BirA enzymes respond to WT BAP by generating AMP. BirA* showed a slightly higher level of AMP production than BirA(WT) in the absence of BAP, which is to be expected since the fast off-rate of bioAMP from this mutant would be followed by a low level of spontaneous bioAMP hydrolysis. The BirA(R118K) mutant is virtually identical to BirA(WT) in terms of AMP production in response to WT BAP peptide.

D. Biotin affinity and bioAMP turnover of BirA analyzed with thin layer chromatography

To estimate the affinities of BirA proteins for biotin, we used TLC to measure the ratio of AMP to bioAMP as a function of biotin concentration. Because this reaction is performed in the absence of BAP, AMP generation reflects the spontaneous hydrolysis of bioAMP. Mutations in the biotin-binding loop of BirA, notably Arg118Gly, reduce the affinity for bioAMP and result in its dissociation from the enzyme active site (Kwon and Beckett, 2000). This assay can be considered a partial reaction compared to what occurs in cells, where biotin would be added to a lysyl group on a proximal protein.

To analyze the affinity of BirA mutants for both biotin and bioAMP, we again utilized [α-^{32}P]-labeled ATP, TLC, and autoradiography. Reactions were prepared and carried out in the same manner as the previously described TLC experiment (section 2.4); however, the concentration of biotin in each reaction ranged from 0 to 250 µM. Each reaction (15 µl) was assembled on ice and contained MBP-BirA (2 µM). Reactions were incubated at 37°C for 1 hr. In duplicate, 2 µl of each reaction was spotted onto TLC plates as described above and analyzed by autoradiography and/or Phosphor Imager (**Fig. 4.5A, C, E**). Total signal in each lane was quantified using the plot lanes tool in ImageJ and the

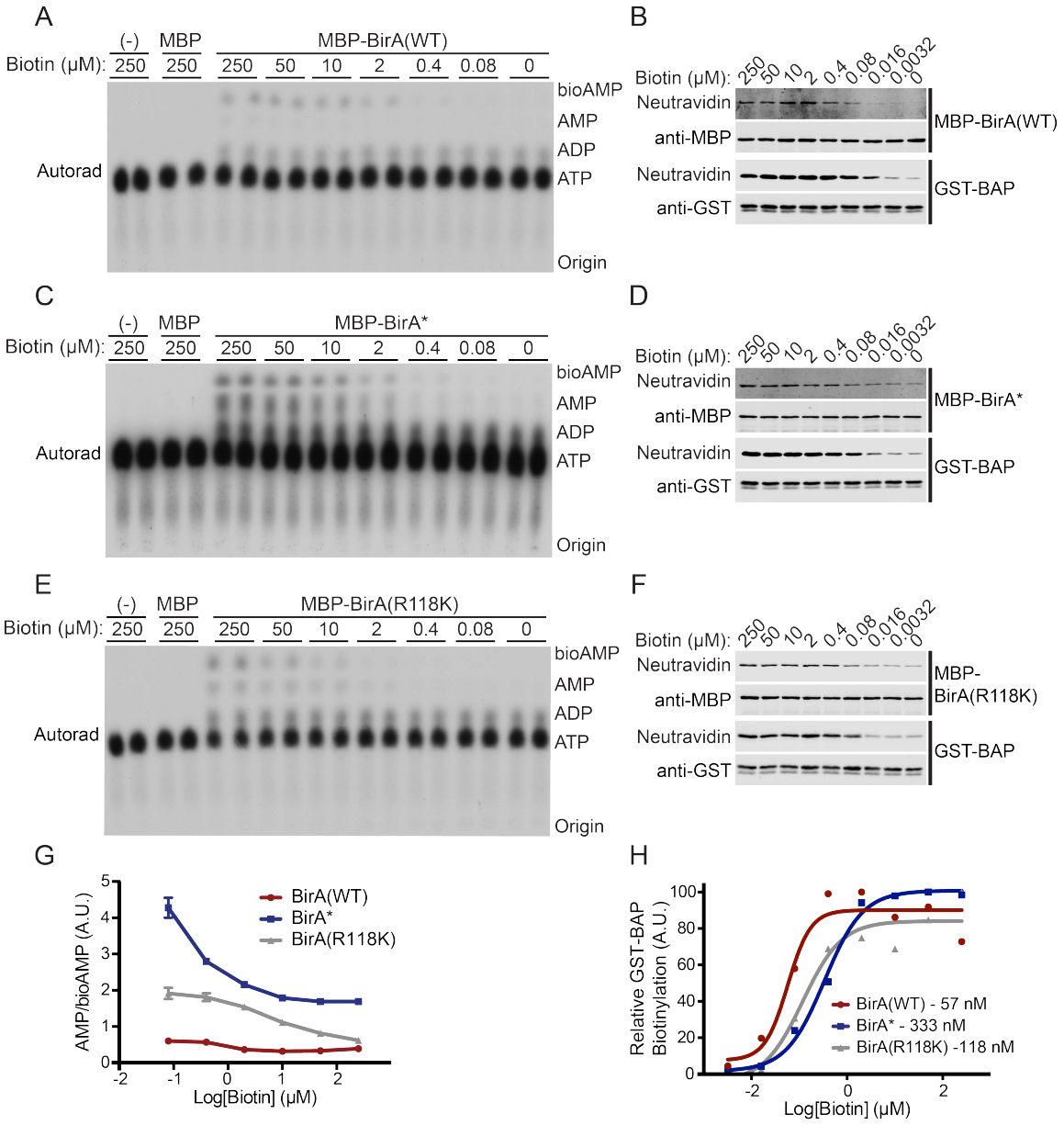


Figure 4.5: Evaluation of BirA biotin affinity and turn-over rate.
TLC analysis of [α-^{32}P]-labeled bioAMP and AMP generation as a function of biotin concentration by **(A)** BirA, **(B)** BirA*, and **(C)** BirA(R118K). Signal for each [α-^{32}P]-labeled species was quantified using ImageJ and the data plotted as the ratio of AMP/bioAMP, a measure of the promiscuous release and hydrolysis of bioAMP to AMP in **(G)**. **(B, D, F)** Biotinylation of BAP measured as a function of biotin concentration using fl-neutra. **(H)** BAP biotinylation is normalized to GST signal and plotted as a function of biotin concentration. Apparent K_M values calculated in GraphPad Prism are shown.

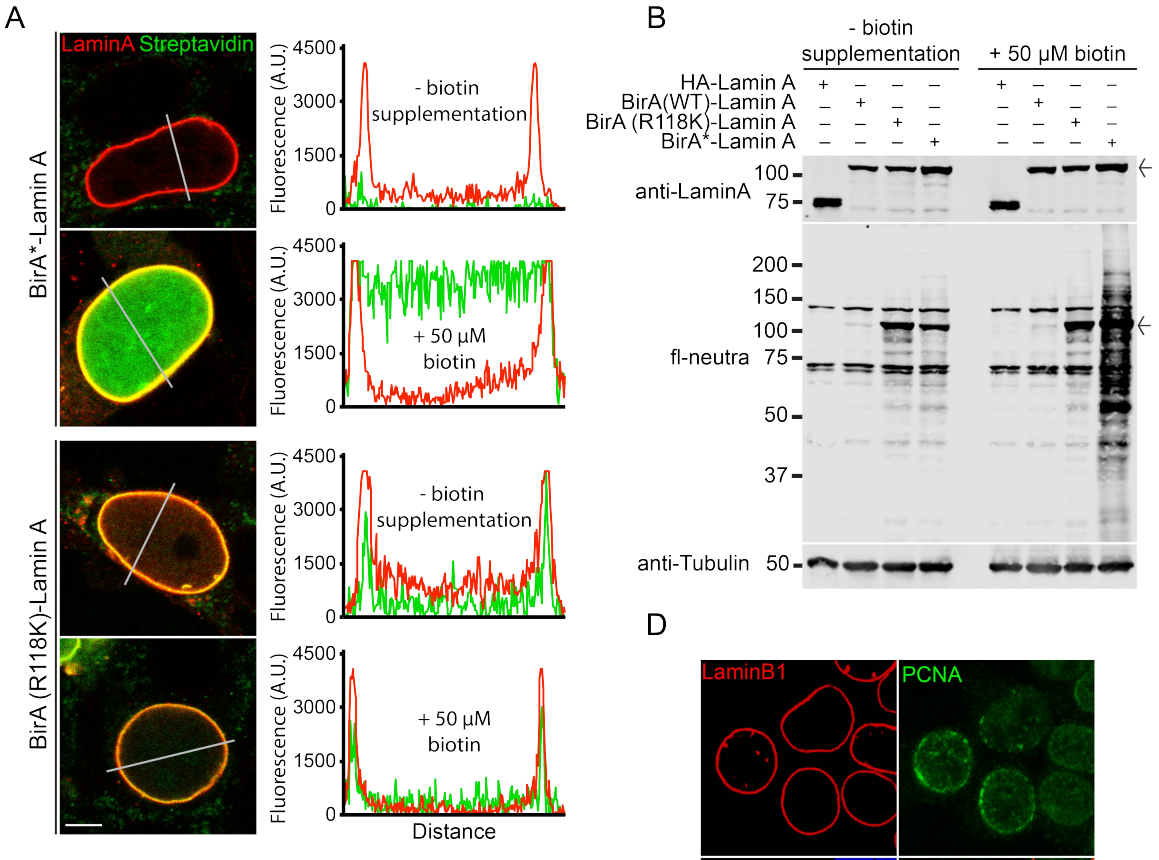
data plotted as a ratio of AMP to bio-AMP (**Fig. 4.5G**). The AMP/bioAMP ratio reflects the basal level of dissociation of bioAMP from BirA and the subsequent hydrolysis of the biotinoyl-AMP linkage prior to TLC. BirA(WT) has an AMP/bioAMP ratio <1 regardless of biotin concentration indicating minimal precocious release of bioAMP. At high biotin concentrations (>50 μM), BirA* readily released bioAMP as indicated by the higher AMP/bioAMP ratio. BirA(Arg118Lys) demonstrated an AMP/bioAMP ratio greater than BirA(WT) (indicating increased precocious bioAMP release) and less than BirA*.

E. Estimation of biotin affinity for BirA enzymes

To determine the approximate K_M of each BirA enzyme for biotin we utilize the GST-BAP substrate. Reactions were prepared on ice in IVBB. Each reaction contained MBP-BirA (50 ng), GST-BAP (1 μg), and ATP (3 mM). Biotin was added to the desired concentration in a 30 μl total reaction volume before placing the sample at 37°C. Each tube was incubated for exactly 20 minutes and then stopped by adding 5X SDS-PAGE sample buffer (6 μl). Half of each reaction was analyzed by SDS-PAGE and immunoblotted with anti-MBP and anti-GST antibodies as well as fl-neutra to detect biotinylation (**Fig. 4.5B, D, F**). Biotinylation of GST-BAP was normalized for protein loading and plotted on a log scale in GraphPad Prism. A non-linear regression was performed, and a sigmoidal fit applied to the data. From this analysis, approximate K_M values were generated (BirA(WT), 57 nM; BirA(R118K), 118 nM; BirA*, 333 nM; **Fig. 4.5H**). These data indicate that BirA(R118K) has a lower affinity for biotin than BirA(WT), but a higher affinity than BirA*.

F. BirA biotinylation visualized by confocal microscopy

The primary application of the characterized BirA mutants is proximity-dependent labeling in cells based on a fusion to a bait protein. For this analysis it is important to confirm that BirA fusion to the bait protein does not interfere with proper localization. We next explored the utility of BirA(R118K) for proximity labeling in cells using Lamin A as the fusion partner. We selected Lamin A for this purpose since BirA*-Lamin A was used in the original BioID method (Roux et al., 2012). The coding region for human Lamin A was PCR amplified and cloned into the pcDNA 3.1 (-) backbone using XhoI



C

	Number of Exclusive Spectra		
	BirA (R118K)-Lamin A	BirA(R118G)-Lamin A	Untransfected
LaminA/C	152	161	0
PCNA	14	0	0
LaminB2	6	27	0
LAP2	2	13	0
TPR	2	6	0

Figure 4.6: Cell-based experiments with BirA-Lamin A fusions.
(A) Confocal microscopy showing the localization (red) and biotinylation (green) of BirA*-Lamin A and BirA(R118K)-Lamin A. HEK293T cells grown in the absence and presence of exogenous biotin (50 µM). Fluorescence intensities were measured across nuclei (white lines) and quantified (middle panels) (Scale bar 5 µm). **(B)** Whole cell lysate from HEK293T cells transfected with HA-Lamin A or BirA-Lamin A fusions with and without biotin supplementation (50 µM). Immunoblot of fl-neutra detection (arrow denotes BirA-Lamin A band). **(C)** Mass spectrometry results obtained from streptavidin pulldown of HEK293T lysate from cells expressing Lamin A-BirA(R118K) and Lamin A-BirA* in the absence of additional biotin supplementation. Shown are selected peptides derived from proteins known to be proximal to Lamin A in cells. A complete list of peptides from this analysis is provided (Table S3). **(D)** Localization of endogenous PCNA to the nuclear lamina by confocal fluorescence microscopy (Scale bar 5 µm).

and AflII restriction sites. Subsequently, the gene encoding BirA(WT) was PCR amplified and cloned into the pcDNA-Lamin A construct using NheI and XhoI restriction sites. This creates an N-terminal BirA-Lamin A fusion. Finally, site-directed mutagenesis was performed to introduce the desired mutations at position 118, in this case R118G and R118K. For cell-based experiments, HEK293T cells were cultured in DMEM/F12 (Gibco; 11320082) supplemented with FBS (5%; Atlanta Biologicals, S11150), Sodium Pyruvate (1%; Gibco, 11360070), MEM Non-Essential Amino Acids (1%; Gibco, 11140076), and Penicillin/Streptomycin (1%; Gibco, 15070063). Transfections were performed using the Fugene6 transfection reagent (Promega, E2691).

Prior to transfection, HEK293T cells were seeded on sterile glass coverslips in 6-well plates ($2.0x10^5$ per well). The next day, cells (~40% confluency) were transfected with BirA-Lamin A fusion plasmids in duplicate wells. The medium for one coverslip of each transfection was supplemented with biotin (50 μM). DMEM/F12 medium contains 14 nM biotin while FBS addition to the medium may contribute up to 10 nM biotin (Dakshinamurti and Chalifour, 1981). This means that medium without biotin supplementation contains a low nanomolar concentration of biotin. After transfection, cells were grown for 24 hrs, and then fixed with formaldehyde (3.75%) in PBS for 10 minutes. Coverslips were washed 3x in PBS, permeabilized with Triton X-100 (0.2%) for 5 minutes and washed again with PBS (3x). Blocking was performed with BSA and FBS (2% each) in PBS for 1 hour. Cells were then stained with anti-Lamin A antibody (Sigma, L1293) for 1 hour, washed in PBS, and incubated with Streptavidin488 (Invitrogen; S11223) and Cy3 anti-rabbit secondary antibody (Jackson ImmunoReseach, 115-165-144) for 1 hour. Coverslips were washed in PBS (3x) and stained with DAPI and mounted on slides using VectaShield (Vector Laboratories; H-1000).

Coverslips were imaged on a Zeiss LSM-880 confocal microscope. The Lamin A signal was used to focus and acquire a slice (1μm thick) through the “middle” of individual nuclei. Image analysis was performed using Zeiss Zen software. Individual nuclei from different experimental conditions but with similar intensities of Lamin A signal were chosen for analysis (**Fig. S4.1**). Traces (indicated by white lines) were drawn across each nucleus and the fluorescence intensity of Lamin A and streptavidin signal was measured

and plotted as a function of distance (**Fig. 4.6A**). Measuring the fluorescence intensities generated from detection of biotin and BirA-Lamin A permits a semi-quantitative comparison of the extent of coincidence. It should be noted that while most Lamin A is stably localized at the nuclear periphery in association with the nuclear lamina, there is also a nucleoplasmic pool of Lamin A. Additionally, since the nuclear lamina is disassembled during mitosis, soluble BirA-Lamin A fusion proteins would have the opportunity to biotinylate non-lamina proteins during part of the cell cycle.

Representative images of nuclei from transfected cells were obtained by confocal microscopy (**Fig. 4.6A**). The BirA(R118K)-Lamin A and BirA*-Lamin A fusions, revealed by staining for Lamin A (red), showed proper localization to the nuclear lamina. Both fusions displayed a low nucleoplasmic signal. Without biotin supplementation, cells expressing BirA* -Lamin A showed a low level of biotinylation by microscopy and blotting, as shown by Roux and co-workers (Roux et al., 2012). Biotinylation by BirA*-Lamin A was greatly enhanced by biotin addition (Fig. 6A, B). BirA(R118K)-Lamin A showed similar levels of biotinylation under both culture conditions (Fig. 6A, B).

G. Identification of lamina-proximal proteins in cells

The next step was to test if the new BirA mutant, BirA(R118K), is capable of proximity labeling in cells. We used BirA(R118K)-Lamin A together with BirA* -Lamin A and untransfected cells as controls. The cell culture component of the experiment was performed without biotin supplementation since the K_M of biotin for the R118K mutant is only two-fold greater than BirA(WT). We acknowledge that this condition might limit the number of protein partners labeled by BirA*-Lamin A given its reduced affinity for biotin (Kwon and Beckett, 2000). BirA*-Lamin A can be viewed as an additional control for the biotin-selective enrichment of proximally labeled proteins. In other words, proteins identified by MS by virtue of enrichment on streptavidin beads in the BirA(R118K)-Lamin A sample cannot be explained by Lamin A binding interactions that are resistant to the stringent buffer conditions if they are not similarly enriched in the BirA*-Lamin A sample.

Plasmids encoding the BirA-Lamin A fusion proteins were each transfected into two 10 cm dishes of HEK293T cells using the Fugene6 transfection reagent. After 24 hours,

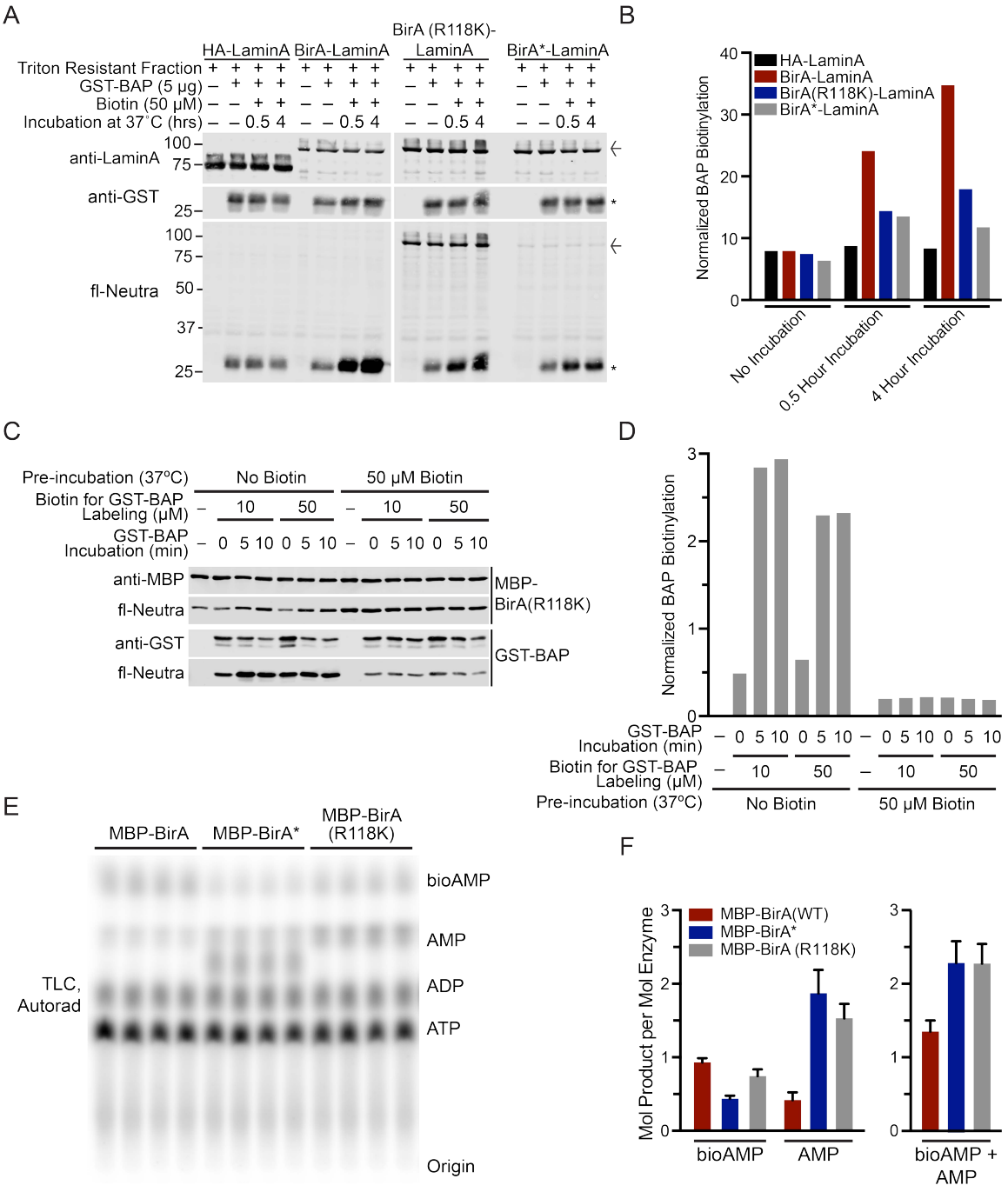


Figure 4.7: Evaluation of BirA(R118K) activity after auto-modification.
(A) Triton resistant fraction from HEK293T cells transfected, in the absence of biotin supplementation, with HA-Lamin A or BirA-Lamin A fusions were harvested and subsequently incubated with biotin for the indicated times. Immunoblot of fl-neutra detection as well (arrow denotes BirA-Lamin A band; asterisk denotes GST-BAP). Signal from fl-neutra on GST-BAP was normalized to GST for each lane and plotted in **(B)**. **(C)** *In vitro* biotinylation of GST-BAP substrate (1 µg) with BirA(R118K) (0.2 µg) with and without a pre-incubation with 50 µM biotin for 2 hours at 37°C. Signal from fl-neutra on GST-BAP was normalized to GST for each lane and

plotted in **(D)**. **(E)** TLC based assay to evaluate production of [α-^{32}P]-labeled bioAMP and AMP by BirA, BirA* and BirA(R118K) after 90 minutes of *in vitro* biotinylation in the presence 50 μM biotin and 50 μM ATP (82.5 nM [α-^{32}P]-ATP) at 37°C. **(F)** Quantification of the mols of product generated per mol of BirA enzyme based on a standard curve quantifying [α-^{32}P]-ATP signal per mol quantified by PhosphorImager (**Fig. S4.2**).

plates were washed with ice cold TBS and scraped into 1.5 ml cold TBS with PMSF (1 mM). Duplicate plates were combined and pelleted (800 x g, 5 min, 4°C). The supernatant was aspirated, and pellets were resuspended in 500 µl lysis buffer (20 mM Tris pH 7.6, 0.65 M NaCl, 0.5% Triton X-100, 0.03% SDS, 2.5 mM EDTA, 2 mM DTT, 1 µM PMSF, 5 µg/mL aprotinin, and 5 µg/ml leupeptin/pepstatin). Tubes were incubated with agitation (4°C for 20 min) followed by tip sonication (10 pulses at 0.4 sec/pulse; Branson Sonifier 250). Lysates were clarified by centrifugation (16,000 x g, 20 min, 4°C). For each condition, 5 µl (1%) was saved as input for the streptavidin pull-down. The lysate was added to streptavidin agarose (150 µl packed; GE Healthcare, 17-5113-01) in a siliconized 1.5 ml tube and incubated with end-over-end rotation (4°C, 4 hrs). After incubation, the supernatant was collected, and beads were washed five times with 500 µl wash buffer (20 mM Tris pH 7.6, 0.25 M NaCl, 0.5% Triton X-100, 0.02% SDS, 1 mM EDTA, 2 mM DTT, 2 µg/mL aprotinin, and 2 µg/ml leupeptin/pepstatin) followed by TBS (500 µl, 5 times). For analysis by SDS-PAGE 1.5 µl (1%) of each sample was removed as input prior to further processing.

Samples, for MS were reduced with DTT (10 mM, 1 hr, RT) followed by alkylation with iodoacetamide (50 mM, 1 hr, RT). Samples were clarified by centrifugation prior to loading of the extract (45%) on the liquid chromatography mass spectrometry system (LC-MS). The LC-MS system consisted of a Thermo Electron Velos Orbitrap ETD mass spectrometer system with a Protana nanospray ion source interfaced to a self-packed 8 cm x 75 µm ID Phenomenex Jupiter 10 µm C18 reversed-phase capillary column. Peptides were eluted from the column with an acetonitrile/0.1 M acetic acid gradient at a flow rate of 0.5 µL/min over 1.3 hours. The nanospray ion source was operated at 2.4 kV and 265C. Instrument settings were as follows: 1 MS scan (FT, 60K resolution, 1 microscan, AGC 9E5, Max IT 500 ms) followed by 20 MS/MS (IT, 1 microscan, AGC 8E3, Max IT 25 ms) with dynamic exclusion enabled (Repeat count 1, Repeat Duration 30 sec, List Size 400, Exclusion Duration 60 sec). This mode of analysis produced approximately 40,000 MS/MS spectra of ions ranging in abundance over several orders of magnitude. These data were analyzed by database searching using the Sequest search algorithm (10 ppm parent, 1 Da fragments, cys – carbamidomethyl fixed, met - oxidized variable) contained within Proteome Discoverer 1.4.1 against the UniProt Human

database (03/2015, 89,663 entries) followed by loading raw search results into Scaffold v4.4.6 (2 peptides, xcorr - +1>1.8, +2>2.0, +3>2.3, +4>2.5, peptide prophet >60%, protein prophet >90%) (for Table S3 see Oostdyk et al., 2019). Proteins with exclusive spectra identified in the untransfected sample were removed and any protein with 2 or greater exclusive spectra were considered positive identifications (for Table S4 see Oostdyk et al., 2019). All MS and analysis was done in the W.M. Keck Biomedical Mass Spectrometry Laboratory at the University of Virginia.

H. Biotin ligase activity of BirA-Lamin A fusions prepared from mammalian cells. We set out to measure the activity of the BirA-Lamin A fusions by biochemical isolation and in vitro biotinylation of the BAP substrate. Given that a small number of peptides derived from lamina-associated proteins were identified by MS using the BirA (R118K)-Lamin A, we considered it possible that this mutant has a lower activity when expressed in cells. Plasmids encoding HA-LaminA and each BirA-Lamin A fusion were transiently expressed in HEK293T cells by transfection with Fugene6. After 24 hours without biotin supplementation, the cells were washed with ice cold TBS and scraped into 1.5 ml cold TBS with PMSF (1 mM). Cells were pelleted by centrifugation (800 x g, 5 min, 4°C). The supernatant was aspirated, and pellets were resuspended in 220 µl lysis buffer (20 mM Tris pH 7.6, 0.5% Triton X-100, 2.5 mM EDTA, 2 mM DTT, 1 µM PMSF, 5 µg/mL aprotinin, and 5 µg/ml leupeptin/pepstatin). Tubes were incubated with agitation (4°C for 20 min) before clarification by centrifugation (16,000 x g, 20 min, 4°C). The supernatant was aspirated and set aside. The triton resistant pellet was then resuspended in IVBB without biotin (150 ul). The pellets were solubilized by sonication (10 pulses at 0.2 sec/pulse; Branson Sonifier 250) and 35 ul was separated into 4 tubes for each transfection. One tube was placed directly on ice. GST-BAP (5 µg) was added to each of the other tubes. Biotin (50 µM) was added to two tubes before incubation at 37°C for either 30 minutes or 4 hours. Reactions were stopped by the addition of 5X SDS sample buffer. Samples (20%) were analyzed by SDS-PAGE and immunoblot (**Fig. 4.7A**). The biotinylation state of GST-BAP for each condition was quantified by normalization of the fl-neutra signal to anti-GST signal and plotted (**Fig 4.7B**).

I. Test for autoinhibition and quantification of bioAMP production.

To test whether the BirA(R118K) mutant is capable of undergoing auto-modification, potentially via biotinylation of the Lys residue engineered into the biotin binding site, we subjected the recombinant MBP-BirA(R118K) to a preincubation without and with biotin (50 μM) to allow for self-biotinylation (**Fig. 4.7C, D**). Samples were then incubated with BAP substrate in a second reaction (0, 5, 10 min). BAP biotinylation was measured and normalized to the GST signal from blotting. To examine the amount of bioAMP and AMP generated by BirA proteins, we mixed a defined amount of each MBP-BirA protein with biotin and [α-^{32}P]-labeled ATP and examined the bioAMP and AMP production by TLC. To estimate the amount of bioAMP and AMP produced per mole of enzyme, we generated a standard curve by spotting 0.003-0.01 pmol of radiolabeled ATP on a TLC plate, exposing the TLC plate on a Phoshor Imager, and plotting the signal as a function of [α-^{32}P]-ATP concentration (**Fig. S4.2**).

IV. Results and Discussion

A. Mutagenesis of position 118 alters BirA biotinylation capabilities.

The mutagenesis of position 118 in BirA (shown in **Fig. 4.2A**) revealed that changing the residue at this position can either enhance or reduce auto-biotinylation of the BirA protein. BirA(WT), with an arginine at position 118, displayed low-level auto-biotinylation, a result of the sequence-specific characteristics of the enzyme and therefore low promiscuity. Substitution with an acidic residue such as aspartate or glutamate resulted in an enzyme which appeared to be catalytically dead, while glycine (the residue in BirA*), lysine, and tryptophan greatly increased native biotinylation levels. Separation of the MBP tag from the BirA enzyme (**Fig. 4.2C**) demonstrated there also were differences in auto-biotinylation of the two species. The high levels of self-modification by BirA(R118K) were almost entirely restricted to BirA, which may partially be a result of biotinylation of the new lysine residue. This high level of specificity, as well as the chemical similarity of lysine to arginine (in BirA(WT)), led us to use BirA(R118K) as our model mutant.

B. BirA(R118K) has properties that are similar to the WT enzyme and mediates proximity labeling in cells

The bead-based assay designed to measure promiscuous self- and trans- biotinylation by BirA mutants underscores the high levels of self-biotinylation by BirA(R118K) (**Fig 4.3C**). Interestingly, BirA* also demonstrated higher self- than trans-biotinylation. This result is surprising and suggests that bioAMP release from BirA* is more effective for self-labeling than trans-modification of MBP in this assay. In contrast, BirA(WT) actually demonstrated greater trans-biotinylation activity than either BirA* or BirA(R118K) suggesting that in this assay BirA(WT) can release bioAMP, which mediates a low level of promiscuous biotinylation.

Analysis of the ability of each BirA mutant to recognize and label WT and mutant BAP sequences demonstrates BirA(WT), BirA*, and BirA(R118K) can each biotinylate the WT BAP (**Fig. 4.4B**). These data suggest that BirA with mutations in position 118 of the biotin-binding loop maintains its recognition and specific biotinylation of a BAP sequence. Biotinylation of the mutant BAP sequence (promiscuous labeling), required a substrate concentration 200-fold greater than WT BAP. This requirement likely reflects low-level bioAMP release from the BirA enzymes and labeling of GST (**Fig. 4.4C**). Consistent with the results of the bead assay (**Fig. 3**) BirA(WT) is capable of a low level of promiscuous biotinylation. Examining the actual production and turnover of bioAMP by TLC supports the ability of each tested BirA to generate bioAMP in the absence of substrate and utilize it for biotinylation when substrate is present (**Fig. 4.4D, E**).

Examination of AMP/bioAMP ratios and biotinylation of BAP, each as a function of biotin concentration, revealed BirA(R118K) has characteristics intermediate to BirA(WT) and BirA*. With both the BirA(R118K) and BirA(WT), the AMP/bioAMP ratio appears to saturate at 50 μM biotin, whereas BirA* did not show signs of saturation at 250 μM biotin (**Fig. 4.5A, C, E, G**). The apparent K_M for biotinylation of a WT BAP sequence is 57, 118, and 333 nM for BirA(WT), BirA(R118K), and BirA* respectively (**Fig. 4.5B, D, F, H**). Thus, relative to the glycine substitution in BirA*, substituting a lysine at position 118 in BirA is associated with higher affinity biotin binding and an intermediate level of bioAMP release.

C. BirA(R118K) demonstrates restricted labeling in cells

To determine if the BirA(R118K) mutant was capable of proximity labeling in cells, HEK293T cells were transfected with Lamin A fusions of BirA(R118K) and BirA*. The cells were grown for 24 hours with and without biotin supplementation and were subsequently examined by confocal fluorescence microscopy as described in section 2.7. We generated fluorescence intensity scans across the nuclei to help establish the location and extent of biotinylation. In BirA*-Lamin A transfected cells, minimal biotinylation was observed without biotin supplementation (**Fig. 4.6A,** green tracing), but biotinylation was increased dramatically by biotin supplementation (**Fig. 4.6A,** upper panels). BirA(R118K) labeling was mostly restricted to the nuclear lamina region with and without biotin supplementation (**Fig. 4.6A,** lower panels). These data suggest that BirA(R118K) may be suitable for proximity labeling of proteins without the necessity of biotin supplementation in culture.

To further evaluate biotinylation by BirA(R118K) in cells, HEK293T cells were transfected with HA-Lamin A, and BirA(WT), BirA(R118K), and BirA* Lamin A fusions in the presence and absence of biotin supplementation (50 μM). Whole cell lysate was assessed by western blot (**Fig. 4.6B**). Probing with fl-neutra again demonstrated increased biotinylation by BirA(R118K) in the absence of biotin supplementation compared to BirA-WT or BirA*. This biotinylation level was affected minimally by the addition of 50 μM biotin, suggesting retention of proximity labeling.

D. BirA(R118K)-Lamin A mediates proximity labeling

We also determined the extent of proximity labeling by BirA(R118K) in cells by streptavidin pulldown and MS analysis (section 2.8) using BirA(R118K)-Lamin A, BirA*-Lamin A, and untransfected cells (**Fig. 4.6C**) in the absence of biotin supplementation. We eliminated proteins from consideration if peptides from said proteins were identified in the control (no transfection sample). We focused on peptides derived from a few extensively characterized proteins of the nuclear lamina and nuclear pore (**Fig. 4.6C**). We found that BirA(R118K)-Lamin A can mediate proximal biotinylation of select nuclear proteins (Lamin A/C, Lamin B2, LAP2, TPR). These proteins were also identified by proximity labeling in the original BioID study that used BirA*-Lamin A (Roux et al., 2012). We found that BirA(R118K)-Lamin A also enabled

identification of PCNA as a proximal protein based on the 14 exclusive spectra from peptides from this protein. PCNA has functional links to the nuclear lamina (Cobb et al., 2016; Dittmer et al., 2014), and is detected proximal to the nuclear lamina in un-transfected HEK293T cells by confocal microscopy (**Fig. 4.6D**). This staining likely reflects centers of replication in late S-phase where both PCNA and Lamin A are known to associate (Kennedy et al., 2000; Shumaker et al., 2008), but it should be noted the signal could be indicative of centers of DNA damage as well (Hwang et al., 1998).

The number of exclusive spectra based on Lamin A peptides identified (which probably reflect self-labeling) was slightly lower for BirA(R118K)-Lamin A compared to BirA* (152 versus 161) (**Fig. 4.6C**). BirA(R118K)-Lamin A also resulted in smaller numbers of exclusive spectra identified for LaminB2 and LAP2 peptides. These data are consistent with the fact that BirA(R118K) is more similar to BirA(WT) in terms of bioAMP release (**Fig. 4.5G**). The surprising finding was that BirA(R118K)-Lamin A was able to label PCNA, a nuclear protein with links to the nuclear lamina in the context of DNA replication (Dittmer et al., 2014; Kennedy et al., 2000; Shumaker et al., 2008). PCNA was not identified by our BirA*-Lamin A sample, and it has not been identified by proximity labeling using Lamin A-BirA* or Lamin B-BirA* in other published studies (Fu et al., 2015; Roux et al., 2012). Why a reduced rate of biotin release by a BirA mutant is apparently advantageous for labeling PCNA is not clear. A list of all proximity labeled proteins that yielded at least two exclusive spectra in either BirA-Lamin A sample is provided (see Table S4 in Oostdyk et al., 2019). A total of 72 and 58 proteins were identified by BirA(R118K)-Lamin A and BirA*-Lamin A, respectively. The small number of proteins identified by BirA*-Lamin A in this experiment reflects the fact that biotin supplementation in culture was not performed. We found that 46 proteins were identified by both BirA-Lamin A fusions, indicating reasonable overlap under the conditions of this experiment.

E. BirA(R118K) is active in cells but can undergo auto-inhibition *in vitro*

The BirA(WT), BirA(R118K), and BirA(R118G) Lamin A fusions immunoprecipitated from cells were biochemically active, as each was able to biotinylated BAP (**Fig. 4.7A, B**). Not surprisingly, the WT enzyme showed the highest activity. We tested whether the

BirA-(R118K) enzyme can undergo auto-inhibition by preincubation with biotin, and assay for its biotin ligase activity toward BAP in a second incubation. This experiment revealed that when preincubated with a large excess of biotin and ATP, MBP-BirA(R118K) activity towards BAP can be significantly reduced. These data indicate that BirA(R118K) auto-modification is auto-inhibitory. The results from the TLC experiment indicate that comparable molar amounts of bioAMP and AMP are generated by the two mutants (**Fig. 4.7E, F**). Collectively, the data indicate that although automodification of BirA(R118K) certainly can occur, it reduces but does not eliminate activity in mammalian cells. BirA(R118K) and BirA* generated comparable levels of bioAMP and AMP, which under our assay conditions are in the range of 1-2 moles of product per mole of enzyme during a 90 min reaction in vitro (**Fig. 4.7E, F**). Thus, BirA(R118K) remains active in the time scale of a BioID experiment.

BioID with the mutant biotin-protein ligase BirA* has proven to be a powerful method for identifying proximal proteins in cells (Roux et al., 2012)). BioID has important advantages over traditional biochemical isolation methods. The promiscuous activity of BirA* is mediated by the Arg118Gly substitution within the flexible biotin binding loop in the active site of BirA. We explored the effects of other substitutions at the same position and identified a substitution, Arg118Lys, that has biochemical properties that can be considered “intermediate” between the WT and Arg118Gly forms of BirA. Our characterization of BirA (Arg118Lys) incorporated multiple biochemical methods developed by other groups working on BirA (Mehus et al., 2016; Roux et al., 2013). We also provided proof-of-principle that BirA (Arg118Lys), despite having a lower activity than BirA* (Choi-Rhee et al., 2004; Kwon and Beckett, 2000; Roux et al., 2012), mediates proximal labeling in cells. Employing BirA mutants that show a range of biotin affinity and biotinoyl-5’-AMP release, including the Arg118Lys variant described here, might be helpful for the discovery of protein interactions by BioID.

A

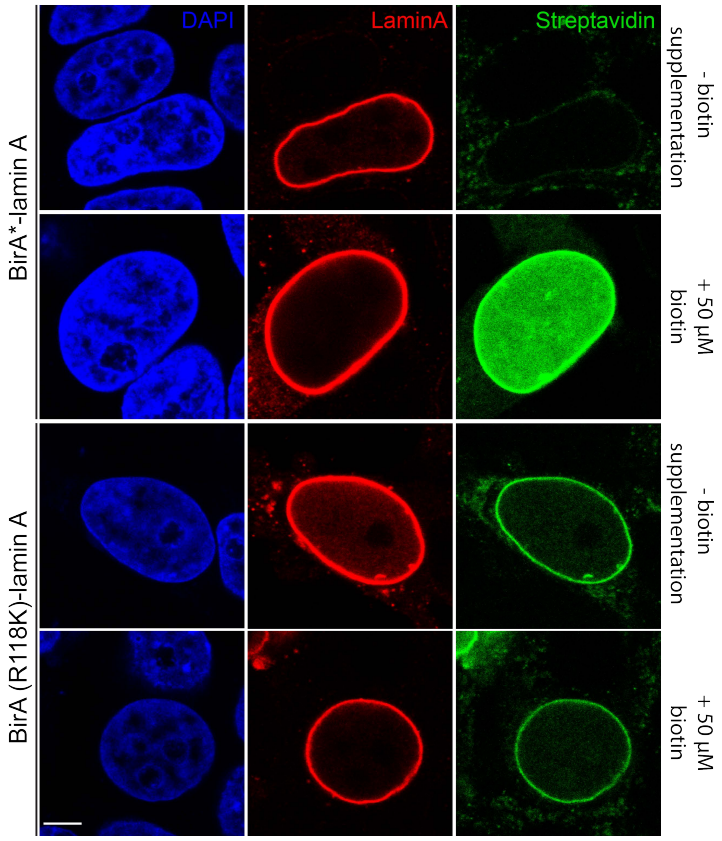


Figure S4.1: Immunofluorescence microscopy of BirA-Lamin A fusions.
(A) Single channel images of HEK293T cells expressing BirA-Lamin A fusions used in **Fig 4.7A**.

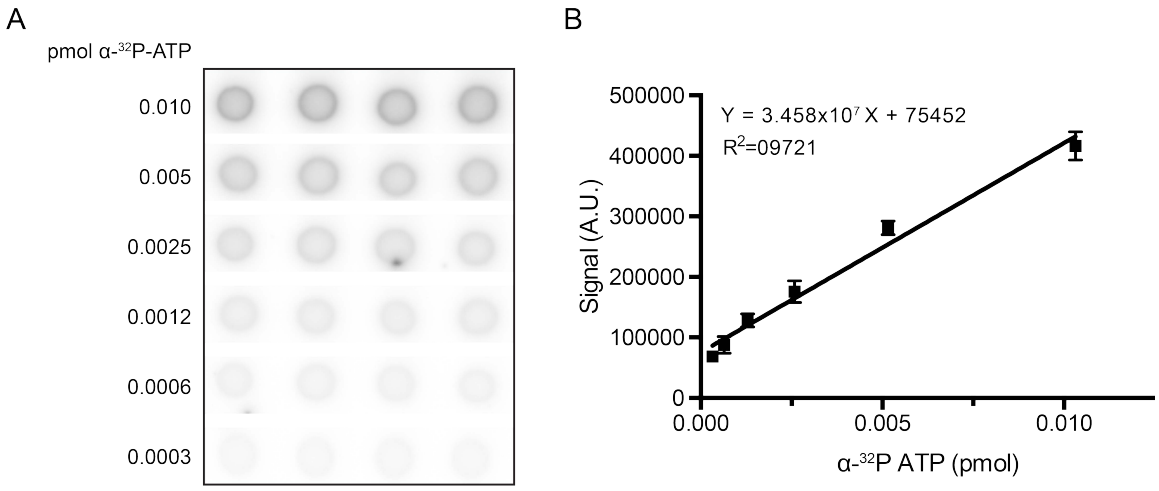


Figure S4.2: [α-^{32}P]-ATP standard curve.
(A) The indicated amounts of [α-^{32}P] ATP was spotted in quadruplicate on a TLC plate and quantified by PhosphorImager. A standard curve for mol of ^{32}P was generated and shown in **(B)**.

Chapter V

Perspectives, Conclusions and Future Directions

As has been discussed throughout this work, the proper function of nuclear transport is a key biological process for maintaining cellular homeostasis, driving developmental programs, and preventing disease states. The classical nuclear transport pathway functions to mediate protein cargoes which contain a nuclear localization signal and is mediated by Imp-α and Imp-β. KPNA7 is the most recently characterized member of the Imp-α family. Previous studies have shown that KPNA7 binds weakly to NLS cargo proteins (Kelley et al., 2010), but can facilitate import of cargoes in vitro (Kimoto et al., 2015). Expression of KPNA7 in multiple species has been shown to be limited to specific tissues and developmental time points (the ovary, oocyte and early stages of embryo development after fertilization) (Hu et al., 2010; Tejomurtula et al., 2009; Wang et al., 2014, 2012). Finally, mutations in KPNA7 have been associated with neurodevelopmental defect and epilepsy (Paciorkowski et al., 2014).

In the course of this study, we have sought to further our knowledge on the functions of KPNA7, from both a biochemical and cellular perspective, as it relates to the regulation of the transport capabilities of the protein, and the impact mutations associated with neurodevelopmental disorder have on these functions. The key findings from the work done in this study include:

i. KPNA7 features differential regulation by the Importin-β binding domain
ii. KPNA7 is expressed in neurons and transport of protein cargoes is reduced by the epilepsy-associated substitution E344Q

In addition to our studies of KPNA7, we also investigated ways to improve the commonly utilized BioID method of discovering protein-protein interactions in cells. BioID utilizes a mutant form of the *E. coli* biotin protein ligase, BirA, termed BirA*, which has an Arg to Gly mutation at position 118 (Roux et al., 2012). This mutation occurs in the biotin binding pocket of the enzyme and functions to reduce affinity of BirA for the intermediate of the biotinylation reaction, biotinoyl-5'-AMP, which allows for its promiscuous release (Cronan, 2005). We performed mutagenic studies of BirA and evaluated the biochemical properties of BirA mutants with each of the 20 possible amino acids residues at position 118 of the enzyme. Through these analyses the key finding of this study was:

iv. BirA(R118K) has distinct, useful biochemical properties which allow for the identification of protein-protein interactions

i. *KPNA7 features differential regulation by the Importin-β binding domain*

In Chapter II of this thesis, we extensively characterized the function of the Importin-β binding domain of KPNA7. The IBB domain is an essential, multifunctional domain in Imp-α proteins which regulates both the binding and release of NLS cargoes, as well as the nuclear import of the receptor via interaction with Imp-β. In yeast, which has a single Imp-α isoform, deletion or mutation of key groups of amino acids within the IBB domain are lethal (Harreman et al., 2003a). Gene duplication events in higher organisms allowed for differential regulation of isoforms (seven in humans) on the basis of expression, NLS cargo specificity and also regulation by the IBB domain (Köhler et al., 1999, 2002; Pumroy and Cingolani, 2015; Pumroy et al., 2015). Specifically, KPNA4 and KPNA6 have been shown to be more weakly auto-inhibited than KPNA2 due to differences of key residues in the backbone of the ARM core and changes in the RxxR motif of the IBB domain (Pumroy et al., 2015).

In our studies, we have determined that KPNA7 is largely relieved from auto-inhibition by its IBB domain. By exchanging the IBB domains between KPNA7 and KPNA2 we have shown this is based on both the primary sequence of the IBB domain, as well as differences in the ARM core of the protein. Sequence analysis points to two potential differences in KPNA7 compared to KPNA2 which could account for the observed effects on auto-inhibition. The first is differences in the composition of the RxxR motif. This motif has been shown to bind to the minor groove of the receptor in the crystal structure of the Imp-α export complex (Matsuura and Stewart, 2004), and, while not crystalized in the auto-inhibited structure of full-length Imp-α (Kobe, 1999), is hypothesized to contribute to auto-inhibition by binding to the minor NLS groove (Pumroy and Cingolani, 2015). In KPNA7, the variable residues in this motif are both glutamines, while in KPNA2, and the majority of Imp-α isoforms both are arginine (see Fig 2.1). The second difference in KPNA7 is the identity of residues in each ARM repeat which have been shown to impact the flexibility of the structure (Pumroy et al., 2015). In KPNA2,

each of these residues is a glycine, while in KPNA7 it has been substituted for asparagine in ARM 4 and aspartic acid in ARM 8 (see Fig 2.S2). Whether these sequence changes completely account for the differences in auto-inhibition we observe between KPNA2 and KPNA7 is unclear, however, our analysis of the IBB swaps is suggestive of a contribution of both the IBB domain and ARM core of the receptor.

Despite the reduced auto-inhibition by the IBB domain of KPNA7, we were still able to observe Imp-β-dependent enhancement of NLS cargo binding. Removal of the IBB domain of Imp-α has been shown to greatly increase the NLS binding affinity of the receptor to low nanomolar levels (Catimel et al., 2001). In contrast, we observe that removal of the KPNA7 IBB domain actually reduces NLS binding. These data suggest that there is a contribution of the IBB domain and its interaction with Imp-β which promote NLS binding, or removal of the IBB domain disrupts the structure of the ARM-core and NLS binding groove. A systematic analysis of KPNA7 and KPNA2 which expands on our IBB swap constructs and individually changes key residues in KPNA7 (such as the RxxR motif and the helix-breaking residues in ARMs 4 and 8) to their identity in KPNA2 and vice versa would be an interesting prospect for further study. This line of study could further define the cause of differences in auto-inhibition of the two proteins, as well as inform the basis of differences in NLS affinity.

Regardless of the exact reasons for the lack of auto-inhibition of KPNA7, the open-state of the receptor has implications for its cellular activity. KPNA7 binding to Imp-β, even without an NLS cargo, results in KPNA7 nuclear translocation. In addition, the formation of the CAS-mediated export complex is inhibited by the KPNA7 open-state. Together, these two characteristics function to promote steady-state nuclear localization of KPNA7 protein in cells.

The literature on nuclear or non-transport functions of Imp-α generally, and KPNA7 specifically, suggests a few potential hypotheses for transport-independent KPNA7 functions that could be the basis of future directions of study. As described in Chapter I (Section III, Subsection F) multiple functions for Imp-α proteins have been identified outside of traditional nuclear transport. These functions, which still rely on binding of NLS or NLS-like sequences, have Imp-α as a sort of molecular chaperone to control

cellular events. A major one of these roles is during mitosis where Imp-α was shown to regulate mitotic spindle assembly by binding to spindle assembly factors such as TPX2, which contains an NLS sequence that binds in the minor NLS binding groove of Imp-α (Carazo-Salas et al., 1999; Giesecke and Stewart, 2010; Gruss et al., 2001). Additionally, after chromosome segregation, Imp-α can similarly regulate the assembly of the nuclear envelope by binding to NLS sequences in lamins and nuclear pore complex proteins (Adam et al., 2008; Forbes et al., 2015). In mouse pre-implantation embryos, KPNA7 is strongly nuclear in localization (Hu et al., 2010). During cell division in the meiosis II oocyte, localizes to the mitotic spindle, but not the chromatin (Hu et al., 2010). Studies in mouse, porcine and bovine embryos have demonstrated loss of KPNA7 results in dysregulation of cell-division during early embryo development and eventual cell-cycle arrest (Hu et al., 2010; Tejomurtula et al., 2009; Wang et al., 2012). In pancreatic cancer cells with high KPNA7 expression, knockdown of KPNA7 results in G1 cell-cycle arrest, defects in mitotic spindle formation, abnormal nuclear morphology and alterations in lamin protein levels (Laurila et al., 2014; Vuorinen et al., 2018). Together these data demonstrate clear, potential functions of KPNA7 outside of traditional nuclear import and within the nucleus. The open-state may make KPNA7 better suited for these functions compared to other Imp-α isoforms.

One alternative explanation for the function of the open-state of KPNA7 is that it may function to simplify the nuclear import pathway to facilitate rapid transport of key, potentially low-affinity protein cargoes. Reduced auto-inhibition in KPNA4 and KPNA6 has been described that allows a unique bipartite NLS sequence in the Influenza A virus polymerase subunit PB2 to bind these two Imp-α isoforms in the absence of Imp-β (Pumroy et al., 2015). It was proposed that this is a mechanism to reduce the trimolecular nuclear import reaction to a pseudo-bimolecular interaction that gives the viral protein a kinetic advantage over traditional NLS-containing proteins. The HIV protein Vpr has also been shown to partially overcome auto-inhibition, and facilitate homodimerization of Imp-α proteins, which similarly gives it a competitive advantage in utilizing the transport machinery (Miyatake et al., 2016). Finally, the apicomplexan parasites *P. falciparum* and *T. gondii* feature a single Imp-α isoform with reduced autoinhibition hypothesized to facilitate rapid nuclear transport and confer an advantage to fast-growing cells like

parasites (Bhatti and Sullivan, 2005; Dey and Patankar, 2018). Thus, there is ample evidence that reduction in auto-inhibition can function to increase transport in some contexts. We contend that the open-state of KPNA7 could allow it to facilitate rapid import of key protein cargoes during early embryogenesis. Parthenogenic studies with a closed, autoinhibited form of KPNA7, such our IBB swap construct, could be used to investigate the necessity of the open-state for KPNA7 function in early development in the context of both transport and non-transport capacities. In addition, another potential function of KPNA7 whereby the open-state may promote dimerization with other Imp-α isoforms and regulate transport is described in the Appendix.

ii. *KPNA7 is expressed in neurons and transport of neuronal protein cargoes is reduced by the epilepsy-associated substitution E344Q*

The focus of Chapter III of this work was the characterization of mutations in KPNA7 associated with a neurodevelopmental disorder. Two mutations in KPNA7 were previously identified that result in amino acid substitutions in the seventh ARM repeat, part of the minor NLS binding groove. E344 is conserved across all seven human Imp-α isoforms and has been shown to interact with a conserved arginine in ARM 6 (Chang et al., 2013; Fontes et al., 2000), as well as directly with some specific NLS sequences (Miyatake et al., 2016). In initial studies, we determined the E344Q substitution in KPNA7 reduces binding to both monopartite and bipartite NLS sequences. NLS binding affinity has been shown to be correlated with Imp-α mediated nuclear import (Fanara et al., 2000). Indeed, we observed a significant decrease in the nuclear import of the SV40 NLS by KPNA7(E344Q). This reduction in NLS binding and transport by KPNA7(E344Q), in addition to our identification of induction of KPNA7 expression in neurons, served as the basis for our experiment to identify neuronal KPNA7 interacting proteins. We utilized SILAC-based mass spectrometry and identified numerous proteins with enrichment factors suggestive of KPNA7 interaction. We specifically characterized two heterogeneous ribonuclear proteins, hnRNP R and hnRNP U. KPNA7 binding to both hnRNP R and hnRNP U from neurons, as well as the validated NLS sequences from each protein, was reduced by the E344Q substitution. Interestingly, the hnRNP R NLS is bipartite while the hnRNP U NLS is monopartite, further demonstrating that the E344Q

substitution effects binding to both classes of NLS. While the P339A substitution is also proximal to the minor NLS binding groove, the proline at position 339 is not directly conserved with other Imp-α isoforms, but instead shifted 2 residues toward the third alpha helix of ARM 7 (see Fig S3.2A). NLS binding by KPNA7(P339A) was similar to, or slightly greater than, the wild type protein for each NLS sequence we tested. KPNA7(P339A) mediated nuclear import was similarly unaffected or slightly increased.

There are multiple implications for a reduced function form of KPNA7. As discussed in Chapter II, KPNA7 adopts an open-state, and strong nuclear localization in cells. We evaluated KPNA7(E344Q) binding to Imp-β, as well as the efficiency of CAS-mediated nuclear export and observed that the open-state of KPNA7 is retained. From our data in Chapter II, we hypothesize there may be nuclear, and non-transport functions for KPNA7 protein. Our discovery of hnRNP R and hnRNP U as KPNA7 interacting proteins additionally suggests they could be potential targets for KPNA7 binding in a nuclear context. Mutations in hnRNP R and hnRNP U are both linked to severe neurodevelopmental defect similar to that observed for KPNA7 (Bramswig et al., 2017; Duijkers et al., 2019). Each hnRNP R and hnRNP U have neuronal specific functions which have been described (Dombert et al., 2014; Liu et al., 2018; Mizutani et al., 2000). The *C. elegans* homolog of hnRNP U is critical for proper expression of the potassium channel SLO-2, a protein with strong links to epileptic disorder (Liu et al., 2018). Mutations in hnRNP U implicated in neurodevelopmental disease primarily result in frameshift or early termination events that would result in complete loss of hnRNP U function (Bramswig et al., 2017). hnRNP R has been shown to have presynaptic localization in motoneurons and promote axon growth via interaction with β-actin mRNA and SMN (Dombert et al., 2014; Rossoll et al., 2003). The bipartite NLS in hnRNP R is directly proximal to the RGG box in the protein which mediates RNA binding, and has also been shown to be required for interaction with SMN (Rossoll et al., 2003). The hnRNP R mutations associated with neurodevelopmental disease occur primarily in this region of the protein (Duijkers et al., 2019). The close proximity of the RGG box and the NLS sequence suggest that RGG box mediated interactions and may be mutually exclusive. While we did not investigate KPNA7 expression in motoneurons, it is conceivable that KPNA7 binding to hnRNP R in a nuclear context could be used to

regulate the interaction of hnRNP R with other proteins. A loss of this regulatory control would be a potential implication of the KPNA7 mutant and could be an interesting area for further study.

We characterized the interaction of KPNA7 with hnRNP R and hnRNP U but it is unlikely they are the only true KPNA7 interacting proteins from those identified in our MS study. Our analysis of the most highly enriched proteins in the KPNA7 sample revealed a bias for proteins with gene ontologies related to RNA processing. Enrichment in proteins with these functions was observed in both of the other studies which have identified KPNA7 interacting proteins in human cells (Kimoto et al., 2015; Vuorinen et al., 2017). In addition to the two hnRNPs, we suggest that KPNA7 may bind these proteins in the nucleus of neurons and regulate their function.

In Chapter III of this study, we also identified induction of *KPNA7* expression during the neuronal differentiation of human iPSC-derived neural progenitors cells and evidence that KPNA7 plays a role in a program of Imp-α isoform switching in neurogenesis. During this differentiation, there is a significant decrease in expression of *KPNA2* which corresponds with an increase in *KPNA7* expression. In addition, expression of *KPNA5* shows small increases. These data contrast with studies in mouse embryonic stem cells which showed also showed the decrease in expression of *Kpna2*, but found a significant, essential induction of *Kpna1* (Yasuhara et al., 2006). These differences may reflect different requirements of Imp-α isoforms in different species, or differences in the specific developmental time points investigated.

While the primary focus of our characterization of the epilepsy-associated mutations in KPNA7 was on transport function, in the course of our investigation of these mutations we identified a binding site for the transcriptional insulator CTCF in the KPNA7 gene that overlays the c.1030-G>C transversion. We determined that CTCF binds to this sequence in exon 7 and that there is CTCF occupancy in stem cells. During differentiation to neurons, this peak is lost and may contribute to the increase in *KPNA7* expression observed. While CTCF peaks are associated with some other Imp-α genes, we did not identify peaks which directly correlate with expression changes during differentiation. The c.1030G>C transversion significantly reduces CTCF binding affinity

for the motif in exon 7 and is predicted to abolish the regulatory control of *KPNA7* gene expression by CTCF. One mechanism of regulation of CTCF binding to DNA is via CpG methylation (Renda et al., 2007). We interrogated exon 7 of KPNA7 for CpG islands and potential CpG methylation sites in the CTCF motif but did not identify the presence of either. Additionally, the G>C transversion does not introduce a new CpG site.

The second epilepsy associated mutation, C.1015C>G is also in exon 7 and proximal to, but not included in the identified CTCF binding motif. While we did not directly interrogate the effect of this mutation on CTCF binding to exon 7, the possibility remains that it could also affect CTCF mediated regulation of KPNA7. If this were the case, CTCF regulatory control of *KPNA7* expression would be absent from both alleles. We have observed precise control of both induction and repression of *KPNA7* expression in neurons and embryological development. It stands to reason that loss of control of expression of both a functional (P339A) and non-functional (E344Q) protein could drive alterations in nuclear transport which function to disrupt key developmental programs. This is a key direction for future study which could lead to a better understanding of how changes in KPNA7 function promote neurodevelopmental disease.

iii. *BirA(R118K) has distinct, useful biochemical properties which allow for identification of protein-protein interactions*

Chapter IV of this study focused on characterizing mutant forms of the *E. coli* biotin protein ligase BirA in an effort to identify a mutant with biochemical characteristics preferable for identifying protein-protein interactions in cells. While our initial goal in this study was to utilize a newly identified BirA mutant for our interrogation of KPNA7 interacting proteins in neurons in Chapter III, we did not ultimately utilize the newly characterized BirA mutant, BirA(R118K), for that experiment. Nonetheless, our investigation into the biochemical functions of BirA mutants at position 118 of the protein furthers the general knowledge of BirA biochemistry and identified characteristics of BirA(R118K) which may prove useful in certain experimental contexts.

The substitution of arginine for lysine at position 118 of BirA resulted in a protein with slightly reduced affinity for biotin, and the activated biotin intermediate bioAMP,

compared to wild-type BirA, but not to the degree of BirA*. Promiscuous release of bioAMP is the feature of BirA* which allows for proximity labeling of proteins. While BirA* is reported to have an effective labeling radius of 10 nm in cells (Kim et al., 2014), in our hands, and those of some of our colleagues, high levels of background, non-proximal biotinylation by BirA* was problematic for identifying true PPIs. The R118K substitution resulted in a BirA enzyme which could facilitate low levels of proximity labeling in cells and was able to identify an interaction of LaminA with PCNA, something which BirA* did not. While we do not claim that BirA(R118K) is superior to BirA* for identifying PPIs in cells, we suggest that in some applications it may be useful to utilize both forms of the enzyme. Further investigation of BirA(R118K) suggested that the enzymatic function of the protein is inhibited by multiple rounds of biotinylation, presumably due to auto-modification of the newly introduced lysine residue in the active site. This functions to reduce the total amount of biotinylation the enzyme can facilitate in cells. An interesting direction for further study would be to investigate the use of BirA(R118K) as a self-modifying protein tag, whereby a protein of interest could be modified with biotin to permit highly specific and high-affinity immunoprecipitation.

In conclusion, this study has characterized novel regulation of the nuclear transport factor KPNA7, particularly with respect to the function of its Imp-β binding domain. These differences result in an Imp-α isoform which is not restricted by auto-inhibition but can bind both Imp-β and NLS sequences independently. These characteristics drive KPNA7 to a nuclear localization and suggest a nuclear function for KPNA7. We have identified neuronal expression of KPNA7 and characterized a mutation in KPNA7 which reduces binding and transport of NLS containing proteins, including those specifically bound in a neuronal setting. Our data suggest neuronal functions of KPNA7 which could be both via NLS binding and transport, and potentially nuclear binding of NLS containing proteins.

Appendix

Selected Additional Studies

Over the course of the studies described in this thesis, other topics and experimental directions have been investigated. Here we describe a few selected preliminary data which have the potential to spur interesting discoveries in the future.

KPNA7 Heterodimerization

Previous studies have identified heterodimerization of KPNA7 with other members of the Imp-α family (Kimoto et al., 2015; Miyamoto and Oka, 2016). This has been proposed as a potential secondary mechanism of regulating nuclear transport in addition to auto-inhibition by the IBB domain. Observations have been made in crystallization studies of ΔIBB-Imp-α that two separate Imp-α proteins can form a closed heterodimer with extensive interaction between the NLS grooves of the two proteins (Conti et al., 1998; Miyatake et al., 2015). The in vivo function of heterodimers is unclear, however it has been proposed as a secondary method of inhibiting Imp-α mediated transport, which might be of particular interest in light of our discovery that auto-inhibition by the IBB domain is reduced for KPNA7. In agreement with previous studies, we have observed KPNA7 has the ability to form heterodimers with other Imp-α isoforms as well as homodimers with itself both in in vitro binding assays and co-immunoprecipitation experiments from cells. We expressed the seven human Imp-α isoforms as ^{35}S-methionine-labeled proteins by in vitro transcription and translation and utilized immobilized GST-KPNA7 to evaluated potential dimerization (**Fig A.1A**). We observed some level of binding each Imp-α isoform to GST-KPNA7, with stronger binding observed by KPNA1, KPNA3, and KPNA5, as well as homodimerization by KPNA7 (**Fig A.1B**). To evaluate if this binding is dependent on the NLS binding groove of either Imp-α protein we evaluated KPNA1 binding to KPNA7 in the presence of SV40 NLS peptide (1 μM) (**Fig A.1C**). Addition of the SV40 NLS reduces KPNA1 binding to KPNA7 and suggests that the NLS binding groove of one, or both, of the two Imp-α isoforms is required for dimerization (**Fig A.1D**). By comparing the ability of the IBB domains from KPNA2 and KPNA7 to bind the members of the Imp-α family we observe that IBB7 is able to bind all members of the Imp-α family, while IBB2 is not (**Fig A.1E, F**). The low binding of IBB7 to full-length KPNA7 agrees with our discovery in Chapter II that KPNA7 is in an open-state. This would also suggest that in cells KPNA7 may be

able to bind to other Imp-α isoforms via its free IBB domain. Finally, co-transfection of HA-KPNA7 with T7-tagged Imp-α isoforms in HEK293T cells demonstrated that KPNA7 can bind with KPNA1, KPNA2 and KPNA6, as well as with itself in cells (**Fig A.1A**). Together these data suggest that there may be a regulatory function for KPNA7 in cells which is dependent on its ability to bind to other Imp-α isoforms via its free IBB domain.

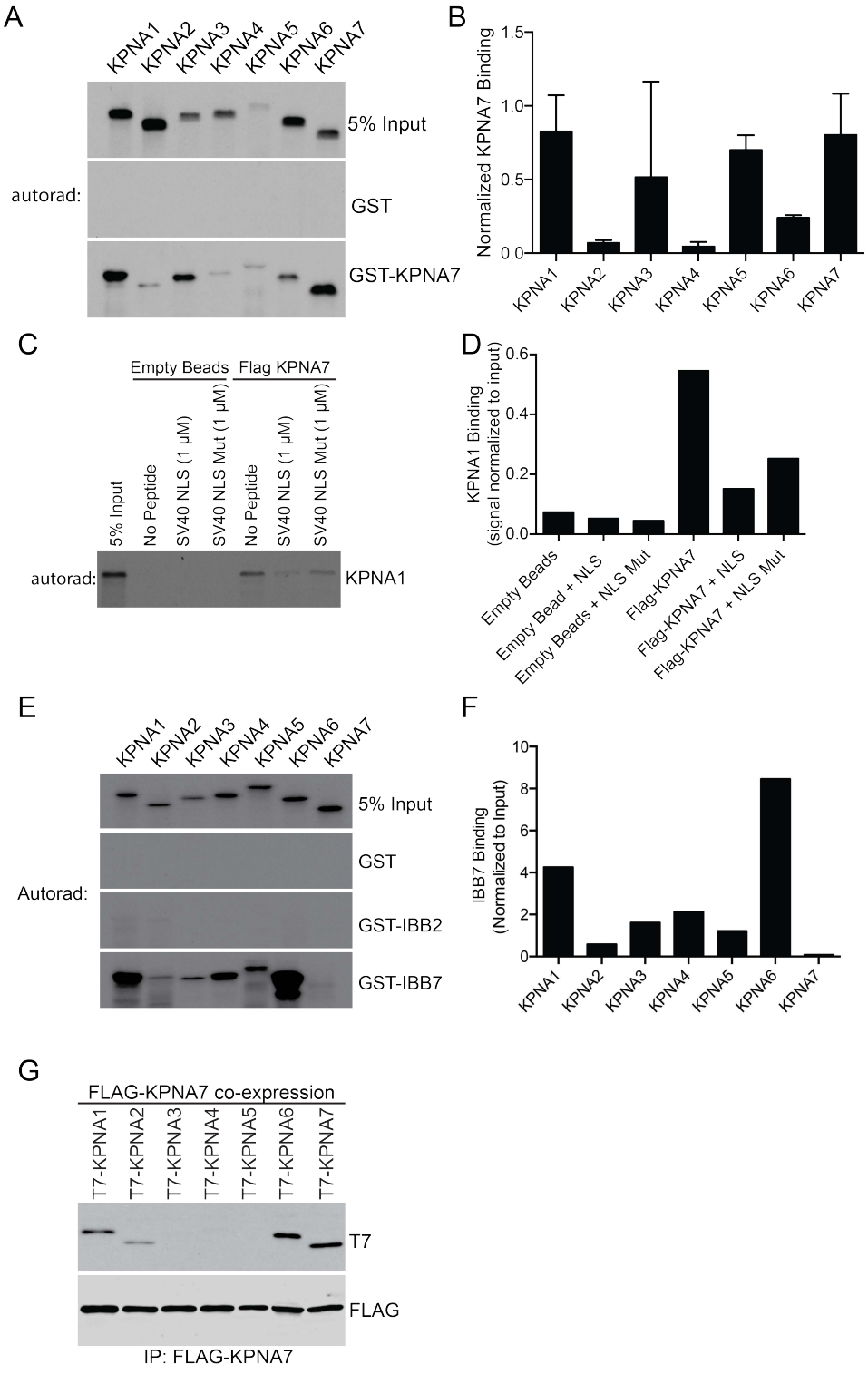

A
KPNA1
KPNA2
KPNA3
KPNA4
KPNA5
KPNA6
KPNA7
5% Input
autorad:
GST
GST-KPNA7
B
Normalized KPNA7 Binding
1.5
1.0
0.5
0.0
KPNA1
KPNA2
KPNA3
KPNA4
KPNA5
KPNA6
KPNA7
C
Empty Beads
Flag KPNA7
5% Input
No Peptide
SV40 NLS (1 μM)
SV40 NLS Mut (1 μM)
No Peptide
SV40 NLS (1 μM)
SV40 NLS Mut (1 μM)
autorad:
KPNA1
D
KPNA1 Binding
(signal normalized to input)
0.6
0.4
0.2
0.0
Empty Beads
Empty Bead + NLS
Empty Beads + NLS Mut
Flag-KPNA7
Flag-KPNA7 + NLS
Flag-KPNA7 + NLS Mut
E
KPNA1
KPNA2
KPNA3
KPNA4
KPNA5
KPNA6
KPNA7
5% Input
Autorad:
GST
GST-IBB2
GST-IBB7
F
IBB7 Binding
(Normalized to Input)
10
8
6
4
2
0
KPNA1
KPNA2
KPNA3
KPNA4
KPNA5
KPNA6
KPNA7
G
FLAG-KPNA7 co-expression
T7-KPNA1
T7-KPNA2
T7-KPNA3
T7-KPNA4
T7-KPNA5
T7-KPNA6
T7-KPNA7
T7
FLAG
IP: FLAG-KPNA7

Figure A1: KPNA7 dimerizes with other Imp-α isoforms.
(A) GST-KPNA7 was immobilized on glutathione resin and used to evaluate binding of *in vitro* translated ^{35}S-methionine-labeled Imp-α isoforms. Samples from duplicate experiments were analyzed by SDS-PAGE, western blot, and autoradiography, quantified in ImageJ and plotted in **(B)**. **(C)** Flag-KPNA7 transfected into HEK293T cells and immunoprecipitated with Flag antibody. KPNA1 was *in vitro* translated as an ^{35}S-methionine-labeled protein, and binding was evaluated ± SV40 NLS peptide (1 μM). Samples were analyzed by SDS-PAGE, western blot, and autoradiography, quantified in ImageJ and plotted in **(D)**. **(E)** The IBB domains from KPNA2 and KPNA7 were immobilized on glutathione beads and used to evaluate binding of *in vitro* translated ^{35}S-methionine-labeled Imp-α isoforms. Samples from duplicate experiments were analyzed by SDS-PAGE, western blot, and autoradiography, quantified in ImageJ and plotted in **(F)**. **(G)** Flag-KPNA7 and T7-tagged Imp-α isoforms were co-transfected into HEK293T cells. Co-immunoprecipitation was performed with Flag antibody and samples were analyzed by SDS-PAGE and western blot.

Expression of KPNA7 in neurons

In Chapter III of this study, we have identified the induction of expression of KPNA7 during neuronal differentiation. While this is a significant ~3.6-fold induction, we were not able to detect KPNA7 protein expression with available antibodies. In the course of this study, we have performed overexpression experiments with KPNA7 and the KPNA7 epilepsy-associated mutants in iPSC-derived neurons. We believe these studies could recapitulate potential increases in expression of the KPNA7 mutants in patients, particularly the E344Q mutant because of the loss of CTCF binding.

To investigate if KPNA7 has an effect on neuronal morphology in human neurons we expressed KPNA7 via lentiviral transduction at low titer in iPSC-derived neurons (**Fig A2A**). These transductions occurred at 1-week of differentiation to reflect the timing of induction of KPNA7 expression. After letting the neuron grow up to 4 or 12 total weeks of differentiation we utilized confocal immunofluorescence microscopy and evaluated the morphological characteristics of the neurons including the number of processes off the soma as well as soma size, two characteristic which have been shown to correlate with maturation of iPSC-derived neurons (Kang et al., 2017). In iPSC-derived neurons, wild-type KPNA7 and both mutants retain their strong nuclear localization (**Fig A2B**). We measured the number of processes of the soma at 4 and 12 weeks of differentiation and noted an increase from 4 to 12 weeks of differentiation for control and KPNA7 expressing cells (**Fig A2C**). At each time point, however, there was no difference between the control or and of the KPNA7 expressing cells. Additionally, measurement of the soma size at 4 weeks of differentiation showed a trend toward reduced soma size in KPNA7(E344Q) expressing cells, but this was not significant (**Fig A2D**). The results of this experiment suggest that overexpression of KPNA7 or individual KPNA7 mutants do not affect neuronal morphology. It should be noted that this is in the context of background, endogenous KPNA7 expression. Follow-up studies could be performed which look at co-expression of the mutants or expression in the context of a KPNA7 knockout cell.

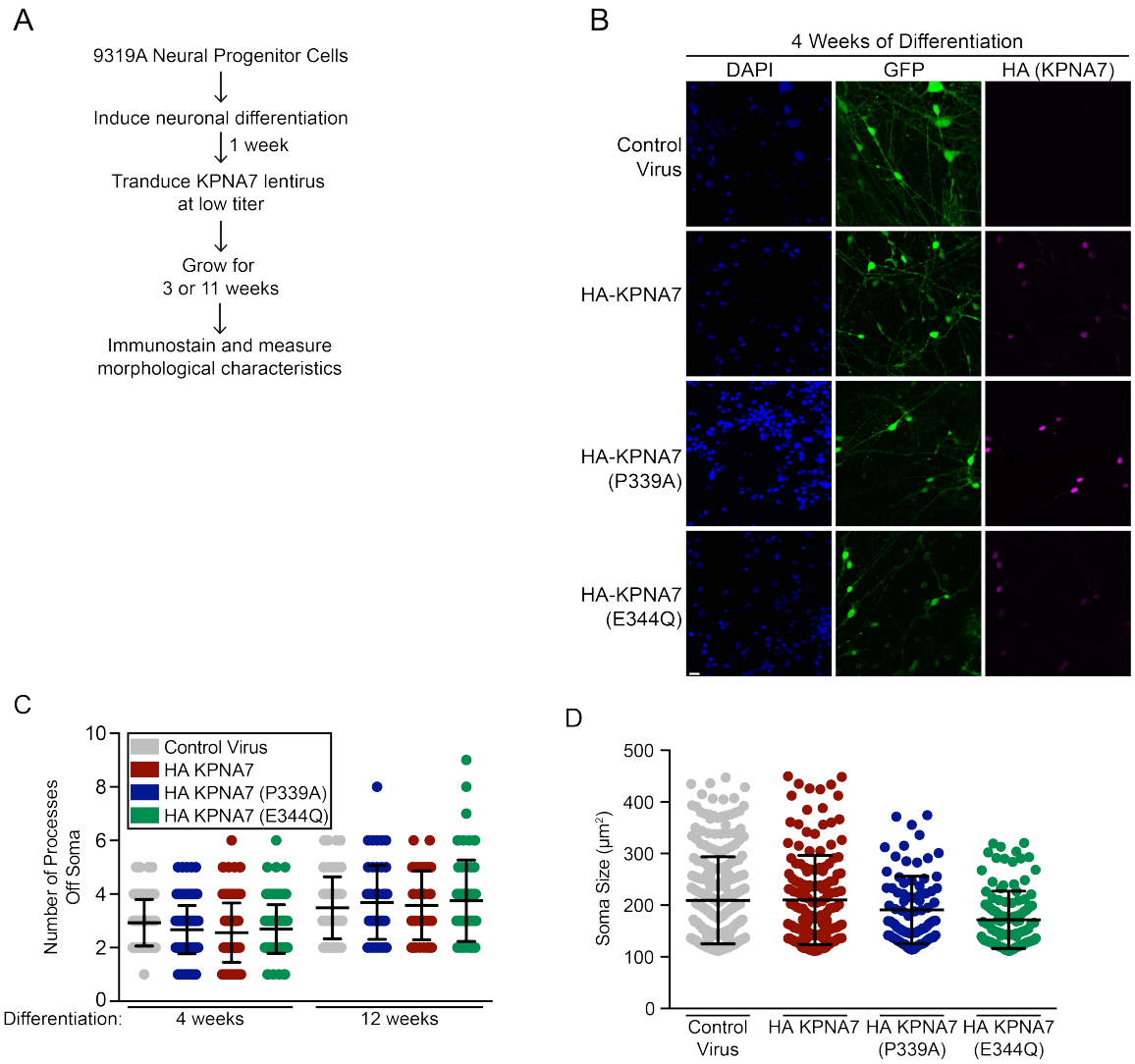


Figure A2: Exogenous expression of KPNA7 and KPNA7 mutants in neurons.
(A) Scheme for expression of KPNA7 and KPNA7 mutants by lentiviral transduction in iPSC-derived neurons. **(B)** Confocal immunofluorescence microscopy of transduced neurons after 4 weeks of differentiation in culture. Cultures were stained using HA antibody to detect transduced HA-KPNA7. **(C)** The number of processes off the soma of transduced neurons at 4 and 12 weeks of differentiation were quantified in ImageJ using the GFP signal with Simple Neurite Tracer program. **(D)** The area of the soma of transduced neurons at 4 weeks of differentiation was quantified in ImageJ using the Simple Neurite Tracer program. No significant differences were observed between the treatments. (Scale: 10 µm)

References

Abdul Razak, A.R., Mau-Soerensen, M., Gabrail, N.Y., Gerecitano, J.F., Shields, A.F., Unger, T.J., Saint-Martin, J.R., Carlson, R., Landesman, Y., McCauley, D., et al. (2016). First-in-Class, First-in-Human Phase I Study of Selinexor, a Selective Inhibitor of Nuclear Export, in Patients With Advanced Solid Tumors. J. Clin. Oncol. *34*, 4142–4150.

Adam, S.A., Sengupta, K., and Goldman, R.D. (2008). Regulation of nuclear lamin polymerization by importin alpha. J. Biol. Chem. *283*, 8462–8468.

Ahluwalia, A., Narula, J., Jones, M.K., Deng, X., and Tarnawski, A.S. (2010). Impaired angiogenesis in aging myocardial microvascular endothelial cells is associated with reduced importin alpha and decreased nuclear transport of HIF1 alpha: mechanistic implications. J. Physiol. Pharmacol. *61*, 133–139.

Ahluwalia, A., Jones, M.K., and Tarnawski, A.S. (2014). Key role of endothelial importin-α in VEGF expression and gastric angiogenesis: novel insight into aging gastropathy. Am. J. Physiol. Liver Physiol. *306*, G338–G345.

Ambron, R.T., and Walters, E.T. (1996). Priming events and retrograde injury signals. Mol. Neurobiol. *13*, 61–79.

Andrade, M.A., Petosa, C., O'Donoghue, S.I., Müller, C.W., and Bork, P. (2001). Comparison of ARM and HEAT protein repeats. J. Mol. Biol. *309*, 1–18.

Bayliss, R., Littlewood, T., and Stewart, M. (2000). Structural Basis for the Interaction between FxFG Nucleoporin Repeats and Importin-β in Nuclear Trafficking. Cell *102*, 99–108.

Beagan, J.A., Duong, M.T., Titus, K.R., Zhou, L., Cao, Z., Ma, J., Lachanski, C. V, Gillis, D.R., and Phillips-Cremins, J.E. (2017). YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. Genome Res. *27*, 1139–1152.

Bednenko, J., Cingolani, G., and Gerace, L. (2003). Importin beta contains a COOH-terminal nucleoporin binding region important for nuclear transport. J. Cell Biol. *162*, 391–401.

Ben-Efraim, I., and Gerace, L. (2001). Gradient of increasing affinity of importin beta for nucleoporins along the pathway of nuclear import. J. Cell Biol. *152*, 411–417.

Ben-Yaakov, K., Dagan, S.Y., Segal-Ruder, Y., Shalem, O., Vuppalanchi, D., Willis, D.E., Yudin, D., Rishal, I., Rother, F., Bader, M., et al. (2012). Axonal transcription factors signal retrogradely in lesioned peripheral nerve. EMBO J. *31*, 1350–1363.

Benarroch, E.E. (2019). Nucleocytoplasmic transport: Mechanisms and involvement in neurodegenerative disease. Neurology *92*, 10.1212/WNL.0000000000007305.

Bhatti, M.M., and Sullivan, W.J. (2005). Histone acetylase GCN5 enters the nucleus via importin-alpha in protozoan parasite Toxoplasma gondii. J. Biol. Chem. *280*, 5902–5908.

Bischoff, F.R., Klebe, C., Kretschmer, J., Wittinghofer, A., and Ponstingl, H. (1994).

RanGAP1 induces GTPase activity of nuclear Ras-related Ran. Proc. Natl. Acad. Sci. U. S. A. *91*, 2587–2591.

Bischoff, F.R., Krebber, H., Smirnova, E., Dong, W., and Ponstingl, H. (1995). Co-activation of RanGTPase and inhibition of GTP dissociation by Ran-GTP binding protein RanBP1. EMBO J. *14*, 705–715.

Black, B.E., Holaska, J.M., Rastinejad, F., and Paschal, B.M. (2001). DNA binding domains in diverse nuclear receptors function as nuclear export signals. Curr. Biol. *11*, 1749–1758.

Bramswig, N.C., Lüdecke, H.-J., Hamdan, F.F., Altmüller, J., Beleggia, F., Elcioglu, N.H., Freyer, C., Gerkes, E.H., Demirkol, Y.K., Knupp, K.G., et al. (2017). Heterozygous HNRNPU variants cause early onset epilepsy and severe intellectual disability. Hum. Genet. *136*, 821–834.

Branon, T.C., Bosch, J.A., Sanchez, A.D., Udeshi, N.D., Svinkina, T., Carr, S.A., Feldman, J.L., Perrimon, N., and Ting, A.Y. (2018). Efficient proximity labeling in living cells and organisms with TurboID. Nat. Biotechnol. *36*, 880–887.

Brennand, K.J., Simone, A., Jou, J., Gelboin-Burkhart, C., Tran, N., Sangar, S., Li, Y., Mu, Y., Chen, G., Yu, D., et al. (2011). Modelling schizophrenia using human induced pluripotent stem cells. Nature *473*, 221–225.

Cappelli, S., Romano, M., and Buratti, E. (2018). Systematic Analysis of Gene Expression Profiles Controlled by hnRNP Q and hnRNP R, Two Closely Related Human RNA Binding Proteins Implicated in mRNA Processing Mechanisms. Front. Mol. Biosci. *5*, 79.

Carazo-Salas, R.E., Guarguaglini, G., Gruss, O.J., Segref, A., Karsenti, E., and Mattaj, I.W. (1999). Generation of GTP-bound Ran by RCC1 is required for chromatin-induced mitotic spindle formation. Nature *400*, 178–181.

Carvill, G.L., Heavin, S.B., Yendle, S.C., McMahon, J.M., O'Roak, B.J., Cook, J., Khan, A., Dorschner, M.O., Weaver, M., Calvert, S., et al. (2013). Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1. Nat. Genet. *45*, 825–830.

Cassany, A., and Gerace, L. (2009). Reconstitution of nuclear import in permeabilized cells. Methods Mol. Biol. *464*, 181–205.

Catimel, B., Teh, T., Fontes, M.R., Jennings, I.G., Jans, D.A., Howlett, G.J., Nice, E.C., and Kobe, B. (2001). Biophysical characterization of interactions involving importin-alpha during nuclear import. J. Biol. Chem. *276*, 34189–34198.

Chakravartty, V., and Cronan, J.E. (2012). Altered regulation of escherichia coli biotin biosynthesis in bira superrepressor mutant strains. J. Bacteriol. *194*, 1113–1126.

Chang, C.W., Couñago, R.M., Williams, S.J., Bodén, M., and Kobe, B. (2013). Distinctive Conformation of Minor Site-Specific Nuclear Localization Signals Bound to Importin-α. Traffic *14*, 1144–1154.

Chatterjee, M., and Paschal, B.M. (2015). Disruption of the ran system by cysteine oxidation of the nucleotide exchange factor RCC1. Mol. Cell. Biol. *35*, 566–581.

Chen, C.-L., and Perrimon, N. (2017). Proximity-dependent labeling methods for proteomic profiling in living cells: WIREs Dev Biol.

Cheng, Y.-S., Seibert, O., Klöting, N., Dietrich, A., Straßburger, K., Fernández-Veledo, S., Vendrell, J.J., Zorzano, A., Blüher, M., Herzig, S., et al. (2015). PPP2R5C Couples Hepatic Glucose and Lipid Homeostasis. PLOS Genet. *11*, e1005561.

Choi-Rhee, E., Schulman, H., and Cronan, J.E. (2004). Promiscuous protein biotinylation by Escherichia coli biotin protein ligase. Protein Sci. *13*, 3043–3050.

Chojnacki, S., Cowley, A., Lee, J., Foix, A., and Lopez, R. (2017). Programmatic access to bioinformatics tools from EMBL-EBI update: 2017. Nucleic Acids Res. *45*, W550–W553.

Chun, Y., Kim, R., and Lee, S. (2016). Centromere Protein (CENP)-W Interacts with Heterogeneous Nuclear Ribonucleoprotein (hnRNP) U and May Contribute to Kinetochore-Microtubule Attachment in Mitotic Cells. PLoS One *11*, e0149127.

Cingolani, G., Petosa, C., Weis, K., and Müller, C.W. (1999a). Structure of importin-beta bound to the IBB domain of importin-alpha. Nature *399*, 221–229.

Cingolani, G., Petosa, C., Weis, K., and Müller, C.W. (1999b). Structure of importin-β bound to the IBB domain of importin-α. Nature *399*, 221–229.

Cobb, A.M., Murray, T. V., Warren, D.T., Liu, Y., and Shanahan, C.M. (2016). Disruption of PCNA-lamin A interactions by prelamin A induces DNA replication fork stalling. Nucleus *7*, 1–22.

Cole, A., Wang, Z., Coyaud, E., Voisin, V., Gronda, M., Jitkova, Y., Mattson, R., Hurren, R., Babovic, S., Maclean, N., et al. (2015). Inhibition of the Mitochondrial Protease ClpP as a Therapeutic Strategy for Human Acute Myeloid Leukemia. Cancer Cell *27*, 864–876.

Comartin, D., Gupta, G.D., Fussner, E., Coyaud, É., Hasegan, M., Archinti, M., Cheung, S.W.T., Pinchev, D., Lawo, S., Raught, B., et al. (2013). CEP120 and SPICE1 Cooperate with CPAP in Centriole Elongation. Curr. Biol. *23*, 1360–1366.

Consortium, E., Allen, A.S., Berkovic, S.F., Cossette, P., Delanty, N., Dlugos, D., Eichler, E.E., Epstein, M.P., Glauser, T., Goldstein, D.B., et al. (2013). De novo mutations in epileptic encephalopathies. Nature *501*, 217–221.

Conti, E., and Kuriyan, J. (2000). Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin α. Structure *8*, 329–338.

Conti, E., Uy, M., Leighton, L., Blobel, G., and Kuriyan, J. (1998). Crystallographic Analysis of the Recognition of a Nuclear Localization Signal by the Nuclear Import Factor Karyopherin α. Cell *94*, 193–204.

Cook, A., Fernandez, E., Lindner, D., Ebert, J., Schlenstedt, G., and Conti, E. (2005). The Structure of the Nuclear Export Receptor Cse1 in Its Cytosolic State Reveals a Closed Conformation Incompatible with Cargo Binding. Mol. Cell *18*, 355–367.

Cook, A., Bono, F., Jinek, M., and Conti, E. (2007). Structural Biology of Nucleocytoplasmic Transport. Annu. Rev. Biochem. *76*, 647–671.

Couzens, A.L., Knight, J.D.R., Kean, M.J., Teo, G., Weiss, A., Dunham, W.H., Lin, Z.-Y., Bagshaw, R.D., Sicheri, F., Pawson, T., et al. (2013). Protein interaction network of the mammalian Hippo pathway reveals mechanisms of kinase-phosphatase interactions. Sci. Signal. *6*, rs15.

Cronan, J.E. (1989). The E. coli bio operon: Transcriptional repression by an essential protein modification enzyme. Cell *58*, 427–429.

Cronan, J.E. (2005). Targeted and proximity-dependent promiscuous protein biotinylation by a mutant Escherichia coli biotin protein ligase. J. Nutr. Biochem. *16*, 416–418.

Dahl, E., Kristiansen, G., Gottlob, K., Klaman, I., Ebner, E., Hinzmann, B., Hermann, K., Pilarsky, C., Dürst, M., Klinkhammer-Schalke, M., et al. (2006). Molecular Profiling of Laser-Microdissected Matched Tumor and Normal Breast Tissue Identifies Karyopherin 2 as a Potential Novel Prognostic Marker in Breast Cancer. Clin. Cancer Res. *12*, 3950–3960.

Dakshinamurti, K., and Chalifour, L.E. (1981). The biotin requirement of HeLa cells. J. Cell. Physiol. *107*, 427–438.

Damelin, M., Silver, P.A., and Corbett, A.H. (2002). Nuclear protein transport. Methods Enzymol. *351*, 587–607.

Dey, V., and Patankar, S. (2018). Molecular basis for the lack of auto-inhibition of Plasmodium falciparum importin α. Biochem. Biophys. Res. Commun. *503*, 1792–1797.

Dingwall, C., and Robbins, J. (1988). The nucleoplasmin nuclear location sequence is larger and more complex than that of SV-40 large T antigen. J. Cell Biol. *107*, 841–849.

Dittmer, T.A., Sahni, N., Kubben, N., Hill, D.E., Vidal, M., Burgess, R.C., Roukos, V., and Misteli, T. (2014). Systematic identification of pathological lamin A interactors. Mol. Biol. Cell *25*, 1493–1510.

Dombert, B., Sivadasan, R., Simon, C.M., Jablonka, S., and Sendtner, M. (2014). Presynaptic Localization of Smn and hnRNP R in Axon Terminals of Embryonic and Postnatal Mouse Motoneurons. PLoS One *9*, e110846.

Dreyfuss, G., Matunis, M.J., Pinol-Roma, S., and Burd, C.G. (1993). hnRNP Proteins and the Biogenesis of mRNA. Annu. Rev. Biochem. *62*, 289–321.

Duijkers, F.A., McDonald, A., Janssens, G.E., Lezzerini, M., Jongejan, A., van Koningsbruggen, S., Leeuwenburgh-Pronk, W.G., Wlodarski, M.W., Moutton, S., Tran-Mau-Them, F., et al. (2019). HNRNPR Variants that Impair Homeobox Gene Expression Drive Developmental Disorders in Humans. Am. J. Hum. Genet. *104*, 1040–1059.

Eggert, M., Michel, J., Schneider, S., Bornfleth, H., Baniahmad, A., Fackelmayer, F.O., Schmidt, S., and Renkawitz, R. (1997). The glucocorticoid receptor is associated with the RNA-binding nuclear matrix protein hnRNP U. J. Biol. Chem. *272*, 28471–28478.

Fagerlund, R., Kinnunen, L., Köhler, M., Julkunen, I., and Melén, K. (2005). NF-κB is transported into the nucleus by importin α3 and importin α4. J. Biol. Chem. *280*, 15942–15951.

Fanara, P., Hodel, M.R., Corbett, A.H., and Hodel, A.E. (2000). Quantitative analysis of nuclear localization signal (NLS)-importin alpha interaction through fluorescence depolarization. Evidence for auto-inhibitory regulation of NLS binding. J. Biol. Chem. *275*, 21218–21223.

Ferreira, P.A. (2019). The coming-of-age of nucleocytoplasmic transport in motor neuron disease and neurodegeneration. Cell. Mol. Life Sci. *76*, 2247–2273.

Firat-Karalar, E.N., and Stearns, T. (2015). Probing mammalian centrosome structure using BioID proximity-dependent biotinylation. Methods Cell Biol. *129*, 153–170.

Firat-Karalar, E.N., Rauniyar, N., Yates, J.R., and Stearns, T. (2014). Proximity Interactions among Centrosome Components Identify Regulators of Centriole Duplication. Curr. Biol. *24*, 664–670.

Fontes, M.R.M., Teh, T., and Kobe, B. (2000). Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin-α. J. Mol. Biol. *297*, 1183–1194.

Fontes, M.R.M., Teh, T., Jans, D., Brinkworth, R.I., and Kobe, B. (2003). Structural basis for the specificity of bipartite nuclear localization sequence binding by importin-alpha. J. Biol. Chem. *278*, 27981–27987.

Forbes, D.J., Travesa, A., Nord, M.S., and Bernis, C. (2015). Nuclear transport factors: global regulation of mitosis. Curr. Opin. Cell Biol. *35*, 78–90.

Fornerod, M., Ohno, M., Yoshida, M., and Mattaj, I.W. (1997). CRM1 Is an Export Receptor for Leucine-Rich Nuclear Export Signals. Cell *90*, 1051–1060.

Fredriksson, K., Van Itallie, C.M., Aponte, A., Gucek, M., Tietgens, A.J., and Anderson, J.M. (2015). Proteomic Analysis of Proteins Surrounding Occludin and Claudin-4 Reveals Their Proximity to Signaling and Trafficking Networks. PLoS One *10*, e0117074.

Fu, Y., Lv, P., Yan, G., Fan, H., Cheng, L., Zhang, F., Dang, Y., Wu, H., and Wen, B. (2015). MacroH2A1 associates with nuclear lamina and maintains chromatin architecture in mouse liver cells. Sci. Rep. *5*, 17186.

Fukuda, A., Nakadai, T., Shimada, M., and Hisatake, K. (2009). Heterogeneous nuclear ribonucleoprotein R enhances transcription from the naturally configured c-fos promoter in vitro. J. Biol. Chem. *284*, 23472–23480.

Fukuda, M., Asano, S., Nakamura, T., Adachi, M., Yoshida, M., Yanagida, M., and Nishida, E. (1997). CRM1 is responsible for intracellular transport mediated by the

nuclear export signal. Nature *390*, 308–311.

Gabriel, G., Klingel, K., Otte, A., Thiele, S., Hudjetz, B., Arman-Kalcek, G., Sauter, M., Shmidt, T., Rother, F., Baumgarte, S., et al. (2011). Differential use of importin-α isoforms governs cell tropism and host adaptation of influenza virus. Nat. Commun. *2*, 156.

Geles, K.G., and Adam, S.A. (2001). Germline and developmental roles of the nuclear transport factor importin alpha3 in C. elegans. Development *128*, 1817–1830.

Giesecke, A., and Stewart, M. (2010). Novel binding of the mitotic regulator TPX2 (target protein for Xenopus kinesin-like protein 2) to importin-alpha. J. Biol. Chem. *285*, 17628–17635.

Gilchrist, D., and Rexach, M. (2003). Molecular basis for the rapid dissociation of nuclear localization signals from karyopherin alpha in the nucleoplasm. J. Biol. Chem. *278*, 51937–51949.

Gilchrist, D., Mykytka, B., and Rexach, M. (2002). Accelerating the rate of disassembly of karyopherin.cargo complexes. J. Biol. Chem. *277*, 18161–18172.

Glinka, M., Herrmann, T., Funk, N., Havlicek, S., Rossoll, W., Winkler, C., and Sendtner, M. (2010). The heterogeneous nuclear ribonucleoprotein-R is necessary for axonal β-actin mRNA translocation in spinal motor neurons. Hum. Mol. Genet. *19*, 1951–1966.

Goldfarb, D.S., Corbett, A.H., Mason, D.A., Harreman, M.T., and Adam, S.A. (2004). Importin alpha: a multipurpose nuclear-transport receptor. Trends Cell Biol. *14*, 505–514.

Görlich, D., and Kutay, U. (1999). Transport between the cell nucleus and the cytoplasm. Annu. Rev. Cell Dev. Biol. *15*, 607–660.

Görlich, D., Vogel, F., Mills, A.D., Hartmann, E., and Laskey, R.A. (1995). Distinct functions for the two importin subunits in nuclear protein import. Nature *377*, 246–248.

Görlich, D., Panté, N., Kutay, U., Aebi, U., and Bischoff, F.R. (1996). Identification of different roles for RanGDP and RanGTP in nuclear protein import. EMBO J. *15*, 5584–5594.

Gousias, K., Becker, A.J., Simon, M., and Niehusmann, P. (2012). Nuclear karyopherin a2: a novel biomarker for infiltrative astrocytomas. J. Neurooncol. *109*, 545–553.

Gousias, K., Niehusmann, P., Gielen, G., Simon, M., and Boström, J. (2014). KPNA2 predicts long term survival in patients with anaplastic oligoastrocytomas. J. Clin. Neurosci. *21*, 1719–1724.

Green, N.M. (1975). Avidin. Adv. Protein Chem. *29*, 85–133.

Gruss, O.J., Carazo-Salas, R.E., Schatz, C.A., Guarguaglini, G., Kast, J., Wilm, M., Le Bot, N., Vernos, I., Karsenti, E., and Mattaj, I.W. (2001). Ran induces spindle assembly by reversing the inhibitory effect of importin α on TPX2 activity. Cell *104*, 83–93.

Gupta, G.D., Coyaud, É., Gonçalves, J., Mojarad, B.A., Liu, Y., Wu, Q., Gheiratmand,

L., Comartin, D., Tkach, J.M., Cheung, S.W.T., et al. (2015). A Dynamic Protein Interaction Landscape of the Human Centrosome-Cilium Interface. Cell *163*, 1484–1499.

Gutierrez, H., and Davies, A.M. (2011). Regulation of neural process growth, elaboration and structural plasticity by NF-κB. Trends Neurosci. *34*, 316–325.

Hamdan, F.F., Srour, M., Capo-Chichi, J.-M., Daoud, H., Nassif, C., Patry, L., Massicotte, C., Ambalavanan, A., Spiegelman, D., Diallo, O., et al. (2014). De Novo Mutations in Moderate or Severe Intellectual Disability. PLoS Genet. *10*, e1004772.

Han, S.P., Tang, Y.H., and Smith, R. (2010). Functional diversity of the hnRNPs: past, present and perspectives. Biochem. J. *430*, 379–392.

Hanz, S., Perlson, E., Willis, D., Zheng, J.-Q., Massarwa, R., Huerta, J.J., Koltzenburg, M., Kohler, M., van-Minnen, J., Twiss, J.L., et al. (2003). Axoplasmic Importins Enable Retrograde Injury Signaling in Lesioned Nerve. Neuron *40*, 1095–1104.

Harel, A., and Forbes, D.J. (2004). Importin beta: conducting a much larger cellular symphony. Mol. Cell *16*, 319–330.

Harreman, M.T., Hodel, M.R., Fanara, P., Hodel, A.E., and Corbett, A.H. (2003a). The Auto-inhibitory Function of Importin Is Essentialin Vivo. J. Biol. Chem. *278*, 5854–5863.

Harreman, M.T., Cohen, P.E., Hodel, M.R., Truscott, G.J., Corbett, A.H., and Hodel, A.E. (2003b). Characterization of the auto-inhibitory sequence within the N-terminal domain of importin alpha. J. Biol. Chem. *278*, 21361–21369.

Hetzer, M., Bilbao-Cortés, D., Walther, T.C., Gruss, O.J., and Mattaj, I.W. (2000). GTP Hydrolysis by Ran Is Required for Nuclear Envelope Assembly. Mol. Cell *5*, 1013–1024.

Hirayama, T., Tarusawa, E., Yoshimura, Y., Galjart, N., and Yagi, T. (2012). CTCF is required for neural development and stochastic expression of clustered Pcdh genes in neurons. Cell Rep. *2*, 345–357.

Hodel, M.R., Corbett, A.H., and Hodel, A.E. (2001). Dissection of a nuclear localization signal. J. Biol. Chem. *276*, 1317–1325.

Hogarth, C.A., Calanni, S., Jans, D.A., and Loveland, K.L. (2006). Importin alpha mRNAs have distinct expression profiles during spermatogenesis. Dev. Dyn. *235*, 253–262.

Hogarth, C.A., Jans, D.A., and Loveland, K.L. (2007). Subcellular distribution of importins correlates with germ cell maturation. Dev. Dyn. *236*, 2311–2320.

Holt, J.E., Ly-Huynh, J.D., Efthymiadis, A., Hime, G.R., Loveland, K.L., and Jans, D.A. (2007). Regulation of Nuclear Import During Differentiation; The IMP alpha Gene Family and Spermatogenesis. Curr. Genomics *8*, 323–334.

Honke, K., and Kotani, N. (2012). Identification of Cell-Surface Molecular Interactions under Living Conditions by Using the Enzyme-Mediated Activation of Radical Sources (EMARS) Method. Sensors *12*, 16037–16045.

Hopkins, C., Gibson, A., Stinchcombe, J., and Futter, C. (2000). Chimeric molecules employing horseradish peroxidase as reporter enzyme for protein localization in the electron microscope. Methods Enzymol. *327*, 35–45.

Hosokawa, K., Nishi, M., Sakamoto, H., Tanaka, Y., and Kawata, M. (2008). Regional distribution of importin subtype mRNA expression in the nervous system: Study of early postnatal and adult mouse. Neuroscience *157*, 864–877.

Hu, J., Wang, F., Yuan, Y., Zhu, X., Wang, Y., Zhang, Y., Kou, Z., Wang, S., and Gao, S. (2010). Novel importin-alpha family member Kpna7 is required for normal fertility and fecundity in the mouse. J. Biol. Chem. *285*, 33113–33122.

Huang, L., Wang, H.-Y., Li, J.-D., Wang, J.-H., Zhou, Y., Luo, R.-Z., Yun, J.-P., Zhang, Y., Jia, W.-H., and Zheng, M. (2013). KPNA2 promotes cell proliferation and tumorigenicity in epithelial ovarian carcinoma through upregulation of c-Myc and downregulation of FOXO3a. Cell Death Dis. *4*, e745–e745.

Hübner, S., Smith, H.M., Hu, W., Chan, C.K., Rihs, H.P., Paschal, B.M., Raikhel, N. V, and Jans, D.A. (1999). Plant importin alpha binds nuclear localization sequences with high affinity and can mediate nuclear import independent of importin beta. J. Biol. Chem. *274*, 22610–22617.

Hung, V., Zou, P., Rhee, H.-W., Udeshi, N.D., Cracan, V., Svinkina, T., Carr, S.A., Mootha, V.K., and Ting, A.Y. (2014). Proteomic Mapping of the Human Mitochondrial Intermembrane Space in Live Cells via Ratiometric APEX Tagging. Mol. Cell *55*, 332–341.

Hwang, B.J., Toering, S., Francke, U., Chu, G., Houtsmuller, A.B., Kanaar, R., and Vermeulen, W. (1998). p48 Activates a UV-damaged-DNA binding factor and is defective in xeroderma pigmentosum group E cells that lack binding activity. Mol. Cell. Biol. *18*, 4391–4399.

Van Itallie, C.M., Aponte, A., Tietgens, A.J., Gucek, M., Fredriksson, K., and Anderson, J.M. (2013). The N and C termini of ZO-1 are surrounded by distinct proteins and functional protein networks. J. Biol. Chem. *288*, 13775–13788.

Van Itallie, C.M., Tietgens, A.J., Aponte, A., Fredriksson, K., Fanning, A.S., Gucek, M., Anderson, J.M., Mareel, M., and van Roy, F. (2014). Biotin ligase tagging identifies proteins proximal to E-cadherin, including lipoma preferred partner, a regulator of epithelial cell-cell and cell-substrate adhesion. J. Cell Sci. *127*, 885–895.

Izaurralde, E., Meyer, S., Teufel, M., Heckel, C., Lührmann, R., and Rautmann, G. (1997). The asymmetric distribution of the constituents of the Ran system is essential for transport into and out of the nucleus. EMBO J. *16*, 6535–6547.

Jeffrey, R.A., Ch'ng, T.H., O'Dell, T.J., and Martin, K.C. (2009). Activity-dependent anchoring of importin alpha at the synapse involves regulated binding to the cytoplasmic tail of the NR1-1a subunit of the NMDA receptor. J. Neurosci. *29*, 15613–15620.

Jensen, J.B., Munksgaard, P.P., Sørensen, C.M., Fristrup, N., Birkenkamp-Demtroder, K., Ulhøi, B.P., Jensen, K.M.-E., Ørntoft, T.F., and Dyrskjøt, L. (2011). High Expression

of Karyopherin-α2 Defines Poor Prognosis in Non–Muscle-Invasive Bladder Cancer and in Patients with Invasive Bladder Cancer Undergoing Radical Cystectomy. Eur. Urol. *59*, 841–848.

Jiang, S., Kotani, N., Ohnishi, T., Miyagawa-Yamguchi, A., Tsuda, M., Yamashita, R., Ishiura, Y., and Honke, K. (2012). A proteomics approach to the cell-surface interactome using the enzyme-mediated activation of radical sources reaction. Proteomics *12*, 54–62.

Jividen, K., Kedzierska, K.Z., Yang, C.-S., Szlachta, K., Ratan, A., and Paschal, B.M. (2018). Genomic analysis of DNA repair genes and androgen signaling in prostate cancer. BMC Cancer *18*, 960.

Kalab, P., and Heald, R. (2008). The RanGTP gradient - a GPS for the mitotic spindle. J. Cell Sci. *121*, 1577–1586.

Kalab, P., Weis, K., and Heald, R. (2002). Visualization of a Ran-GTP gradient in interphase and mitotic Xenopus egg extracts. Science *295*, 2452–2456.

Kalderon, D., Roberts, B.L., Richardson, W.D., and Smith, A.E. (1984). A short amino acid sequence able to specify nuclear location. Cell *39*, 499–509.

Kang, S., Chen, X., Gong, S., Yu, P., Yau, S., Su, Z., Zhou, L., Yu, J., Pan, G., and Shi, L. (2017). Characteristic analyses of a neural differentiation model from iPSC-derived neuron according to morphology, physiology, and global gene expression pattern. Sci. Rep. *7*, 12233.

Kau, T.R., Way, J.C., and Silver, P.A. (2004). Nuclear transport and cancer: from mechanism to intervention. Nat. Rev. Cancer *4*, 106–117.

Kelley, J.B., Talley, A.M., Spencer, A., Gioeli, D., and Paschal, B.M. (2010). Karyopherin alpha7 (KPNA7), a divergent member of the importin alpha family of nuclear import receptors. BMC Cell Biol. *11*, 63.

Kennedy, B.K., Barbie, D.A., Classon, M., Dyson, N., and Harlow, E. (2000). Nuclear organization of DNA replication in primary mammalian cells. Genes Dev. *14*, 2855–2868.

Kiledjian, M., and Dreyfuss, G. (1992). Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. EMBO J. *11*, 2655–2664.

Kim, B.J., and Lee, H. (2008). Caspase-mediated cleavage of importin-α increases its affinity for MCM and downregulates DNA synthesis by interrupting the binding of MCM to chromatin. Biochim. Biophys. Acta - Mol. Cell Res. *1783*, 2287–2293.

Kim, D.I., Birendra, K.C., Zhu, W., Motamedchaboki, K., Doye, V., and Roux, K.J. (2014). Probing nuclear pore complex architecture with proximity-dependent biotinylation. Proc. Natl. Acad. Sci. U. S. A. *111*, E2453-61.

Kim, D.I., Jensen, S.C., Noble, K.A., KC, B., Roux, K.H., Motamedchaboki, K., and Roux, K.J. (2016). An improved smaller biotin ligase for BioID proximity labeling. Mol. Biol. Cell *27*, 1188–1196.

Kim, D.I., Cutler, J.A., Na, C.H., Reckel, S., Renuse, S., Madugundu, A.K., Tahir, R., Goldschmidt, H.L., Reddy, K.L., Huganir, R.L., et al. (2018a). BioSITe: A Method for Direct Detection and Quantitation of Site-Specific Biotinylation. J. Proteome Res. *17*, 759–769.

Kim, H.W., Quan, Z., Kim, Y.-B., Cheong, E., Kim, H.D., Cho, M., Jang, J., Yoo, Y.R., Lee, J.S., Kim, J.H., et al. (2018b). Differential effects on sodium current impairments by distinct SCN1A mutations in GABAergic neurons derived from Dravet syndrome patients. Brain Dev. *40*, 287–298.

Kim, I.S., Kim, D.H., Han, S.M., Chin, M.U., Nam, H.J., Cho, H.P., Choi, S.Y., Song, B.J., Kim, E.R., Bae, Y.S., et al. (2000). Truncated form of importin alpha identified in breast cancer cell inhibits nuclear import of p53. J. Biol. Chem. *275*, 23139–23145.

Kimoto, C., Moriyama, T., Tsujii, A., Igarashi, Y., Obuse, C., Miyamoto, Y., Oka, M., and Yoneda, Y. (2015). Functional characterization of importin α8 as a classical nuclear localization signal receptor. Biochim. Biophys. Acta *1853*, 2676–2683.

Kobe, B. (1999). Autoinhibition by an internal nuclear localization signal revealed by the crystal structure of mammalian importin α. Nat. Struct. Biol. *6*, 388–397.

Köhler, M., Ansieau, S., Prehn, S., Leutz, A., Haller, H., and Hartmann, E. (1997). Cloning of two novel human importin-α subunits and analysis of the expression pattern of the importin-α protein family. FEBS Lett. *417*, 104–108.

Köhler, M., Speck, C., Christiansen, M., Bischoff, F.R., Prehn, S., Haller, H., Görlich, D., and Hartmann, E. (1999). Evidence for Distinct Substrate Specificities of Importin α Family Members in Nuclear Protein Import. Mol. Cell. Biol. *19*, 7782–7791.

Köhler, M., Fiebeler, A., Hartwig, M., Thiel, S., Prehn, S., Kettritz, R., Luft, F., and Hartmann, E. (2002). Differential expression of classical nuclear transport factors during cellular proliferation and differentiation. Cell. Physiol. Biochem. *12*, 335–344.

Koike, M., Sakaki, S., Amano, Y., and Kurosawa, H. (2007). Characterization of embryoid bodies of mouse embryonic stem cells formed under various culture conditions and estimation of differentiation status of such bodies. J. Biosci. Bioeng. *104*, 294–299.

Kosugi, S., Hasebe, M., Matsumura, N., Takashima, H., Miyamoto-Sato, E., Tomita, M., and Yanagawa, H. (2009a). Six classes of nuclear localization signals specific to different binding grooves of importin alpha. J. Biol. Chem. *284*, 478–485.

Kosugi, S., Hasebe, M., Tomita, M., and Yanagawa, H. (2009b). Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. Proc. Natl. Acad. Sci. U. S. A. *106*, 10171–10176.

Kutay, U., Bischoff, F.R., Kostka, S., Kraft, R., and Görlich, D. (1997). Export of Importin α from the Nucleus Is Mediated by a Specific Nuclear Transport Factor. Cell *90*, 1061–1071.

Kwon, K., and Beckett, D. (2000). Function of a conserved sequence motif in biotin holoenzyme synthetases. Protein Sci. *9*, 1530–1539.

Laitman, B.M., Mariani, J.N., Zhang, C., Sawai, S., and John, G.R. (2017). Karyopherin Alpha Proteins Regulate Oligodendrocyte Differentiation. PLoS One *12*, e0170477.

Laurila, E., Savinainen, K., Kuuselo, R., Karhu, R., and Kallioniemi, A. (2009). Characterization of the 7q21-q22 amplicon identifies ARPC1A, a subunit of the Arp2/3 complex, as a regulator of cell migration and invasion in pancreatic cancer. Genes, Chromosom. Cancer *48*, 330–339.

Laurila, E., Vuorinen, E., Savinainen, K., Rauhala, H., and Kallioniemi, A. (2014). KPNA7, a nuclear transport receptor, promotes malignant properties of pancreatic cancer cells in vitro. Exp. Cell Res. *322*, 159–167.

Lee, H., Ueda, M., Miyamoto, Y., Yoneda, Y., Perry, G., Smith, M.A., and Zhu, X. (2006). Aberrant localization of importin alpha1 in hippocampal neurons in Alzheimer disease. Brain Res. *1124*, 1–4.

Lee, H.R., Kim, T.D., Kim, H.J., Jung, Y., Lee, D., Lee, K.H., Kim, D.Y., Woo, K.C., and Kim, K.T. (2015). Heterogeneous ribonucleoprotein R regulates arylalkylamine N-acetyltransferase synthesis via internal ribosomal entry site-mediated translation in a circadian manner. J. Pineal Res. *59*, 518–529.

van de Leemput, J., Boles, N.C., Kiehl, T.R., Corneo, B., Lederman, P., Menon, V., Lee, C., Martinez, R.A., Levi, B.P., Thompson, C.L., et al. (2014). CORTECON: a temporal transcriptome analysis of in vitro human cerebral cortex development from human embryonic stem cells. Neuron *83*, 51–68.

Lefebvre, S., Bürglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M., et al. (1995). Identification and characterization of a spinal muscular atrophy-determining gene. Cell *80*, 155–165.

Leung, S.W., Harreman, M.T., Hodel, M.R., Hodel, A.E., and Corbett, A.H. (2003). Dissection of the karyopherin alpha nuclear localization signal (NLS)-binding groove: functional requirements for NLS binding. J. Biol. Chem. *278*, 41947–41953.

Lever, M.B., Karpova, A., and Kreutz, M.R. (2015). An Importin Code in neuronal transport from synapse-to-nucleus? Front. Mol. Neurosci. *8*, 33.

Li, P., Li, J., Wang, L., and Di, L.-J. (2017). Proximity Labeling of Interacting Proteins: Application of BioID as a Discovery Tool. Proteomics *17*, 1700002.

Li, X.-W., Rees, J.S., Xue, P., Zhang, H., Hamaia, S.W., Sanderson, B., Funk, P.E., Farndale, R.W., Lilley, K.S., Perrett, S., et al. (2014). New insights into the DT40 B cell receptor cluster using a proteomic proximity labeling assay. J. Biol. Chem. *289*, 14434–14447.

Li, Y., Park, K.-E., and Cabot, R.A. (2015). Dynamic changes in nuclear import of a nuclear localisation signal-bearing substrate in 8-cell stage porcine embryos. Reprod. Fertil. Dev. *27*, 385.

Liu, P., Wang, S.J., Wang, Z.-W., and Chen, B. (2018). HRPU-2, a Homolog of Mammalian hnRNP U, Regulates Synaptic Transmission by Controlling the Expression of SLO-2 Potassium Channel in Caenorhabditis elegans. J. Neurosci. *38*, 1073–1084.

Lott, K., and Cingolani, G. (2011). The importin β binding domain as a master regulator of nucleocytoplasmic transport. Biochim. Biophys. Acta - Mol. Cell Res. *1813*, 1578–1592.

Lowe, A.R., Siegel, J.J., Kalab, P., Siu, M., Weis, K., and Liphardt, J.T. (2010). Selectivity mechanism of the nuclear pore complex characterized by single cargo tracking. Nature *467*, 600–603.

Ma, J., and Cao, X. (2006). Regulation of Stat3 nuclear import by importin α5 and importin α7 via two different functional sequence elements. Cell. Signal. *18*, 1117–1126.

Mahipal, A., and Malafa, M. (2016). Importins and exportins as therapeutic targets in cancer. Pharmacol. Ther. *164*, 135–143.

Martell, J.D., Deerinck, T.J., Sancak, Y., Poulos, T.L., Mootha, V.K., Sosinsky, G.E., Ellisman, M.H., and Ting, A.Y. (2012). Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. Nat. Biotechnol. *30*, 1143–1148.

Mason, D.A., Fleming, R.J., and Goldfarb, D.S. (2002). Drosophila melanogaster importin alpha1 and alpha3 can replace importin alpha2 during spermatogenesis but not oogenesis. Genetics *161*, 157–170.

Mason, D.A., Stage, D.E., and Goldfarb, D.S. (2009). Evolution of the metazoan-specific importin alpha gene family. J. Mol. Evol. *68*, 351–365.

Matsuura, Y., and Stewart, M. (2004). Structural basis for the assembly of a nuclear export complex. Nature *432*, 872–877.

Mattaj, I.W., and Englmeier, L. (1998). NUCLEOCYTOPLASMIC TRANSPORT: The Soluble Phase. Annu. Rev. Biochem. *67*, 265–306.

McAllister, H.C., and Coon, M.J. (1966). Further studies on the properties of liver propionyl coenzyme A holocarboxylase synthetase and the specificity of holocarboxylase formation. J. Biol. Chem. *241*, 2855–2861.

Mehus, A.A., Anderson, R.H., and Roux, K.J. (2016). BioID Identification of Lamin-Associated Proteins. Methods Enzymol. *569*, 3–22.

Melchior, F., Paschal, B., Evans, J., and Gerace, L. (1993). Inhibition of nuclear protein import by nonhydrolyzable analogues of GTP and identification of the small GTPase Ran/TC4 as an essential transport factor. J. Cell Biol. *123*, 1649–1659.

Merkenschlager, M., and Nora, E.P. (2016). CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. Annu. Rev. Genomics Hum. Genet. *17*, 17–43.

Michel, N., Majumdar, U.B., Lannigan, J., and McConnell, M.J. (2019). Imaging Flow Cytometry Quantifies Neural Genome Dynamics. Cytom. Part A cyto.a.23783.

Mihalas, B.P., Western, P.S., Loveland, K.L., McLaughlin, E.A., and Holt, J.E. (2015). Changing expression and subcellular distribution of karyopherins during murine oogenesis. Reproduction *150*, 485–496.

Miyamoto, Y., and Oka, M. (2016). Data on dimer formation between importin α

subtypes. Data Br. *7*, 1248–1253.

Miyamoto, Y., Boag, P.R., Hime, G.R., and Loveland, K.L. (2012). Regulated nucleocytoplasmic transport during gametogenesis. Biochim. Biophys. Acta - Gene Regul. Mech. *1819*, 616–630.

Miyamoto, Y., Yamada, K., and Yoneda, Y. (2016). Importin α: a key molecule in nuclear transport and non-transport functions. J. Biochem. *160*, 69–75.

Miyatake, H., Sanjoh, A., Unzai, S., Matsuda, G., Tatsumi, Y., Miyamoto, Y., Dohmae, N., and Aida, Y. (2015). Crystal structure of human importin-a1 (Rch1), revealing a potential autoinhibition mode involving homodimerization. PLoS One *10*, 1–14.

Miyatake, H., Sanjoh, A., Murakami, T., Murakami, H., Matsuda, G., Hagiwara, K., Yokoyama, M., Sato, H., Miyamoto, Y., Dohmae, N., et al. (2016). Molecular Mechanism of HIV-1 Vpr for Binding to Importin-α. J. Mol. Biol. *428*, 2744–2757.

Mizutani, A., Fukuda, M., Ibata, K., Shiraishi, Y., and Mikoshiba, K. (2000). SYNCRIP, a cytoplasmic counterpart of heterogeneous nuclear ribonucleoprotein R, interacts with ubiquitous synaptotagmin isoforms. J. Biol. Chem. *275*, 9823–9831.

Moore, M.S., and Blobel, G. (1993). The GTP-binding protein Ran/TC4 is required for protein import into the nucleus. Nature *365*, 661–663.

Moriyama, T., Nagai, M., Oka, M., Ikawa, M., Okabe, M., and Yoneda, Y. (2011). Targeted disruption of one of the importin α family members leads to female functional incompetence in delivery. FEBS J. *278*, 1561–1572.

Morris, C.P., Baune, B.T., Domschke, K., Arolt, V., Swagell, C.D., Hughes, I.P., Lawford, B.R., McD Young, R., and Voisey, J. (2012). KPNA3 variation is associated with schizophrenia, major depression, opiate dependence and alcohol dependence. Dis. Markers *33*, 163–170.

Mortezavi, A., Hermanns, T., Seifert, H.-H., Baumgartner, M.K., Provenzano, M., Sulser, T., Burger, M., Montani, M., Ikenberg, K., Hofstadter, F., et al. (2011). KPNA2 Expression Is an Independent Adverse Predictor of Biochemical Recurrence after Radical Prostatectomy. Clin. Cancer Res. *17*, 1111–1121.

Mosammaparast, N., and Pemberton, L.F. (2004). Karyopherins: From nuclear-transport mediators to nuclear-function regulators. Trends Cell Biol. *14*, 547–556.

Moseley, G.W., Roth, D.M., DeJesus, M.A., Leyton, D.L., Filmer, R.P., Pouton, C.W., and Jans, D.A. (2007). Dynein Light Chain Association Sequences Can Facilitate Nuclear Protein Import. Mol. Biol. Cell *18*, 3204–3213.

Mourelatos, Z., Abel, L., Yong, J., Kataoka, N., and Dreyfuss, G. (2001). SMN interacts with a novel family of hnRNP and spliceosomal proteins. EMBO J. *20*, 5443–5452.

Nenortas, E., and Beckett, D. (1996). Purification and characterization of intact and truncated forms of the Escherichia coli biotin carboxyl carrier subunit of acetyl-CoA carboxylase. J. Biol. Chem. *271*, 7559–7567.

Ni, L., Llewellyn, R., Kesler, C.T., Kelley, J.B., Spencer, A., Snow, C.J., Shank, L., and Paschal, B.M. (2013). Androgen induces a switch from cytoplasmic retention to nuclear import of the androgen receptor. Mol. Cell. Biol. *33*, 4766–4778.

O'Reilly, A.J., Dacks, J.B., and Field, M.C. (2011). Evolution of the Karyopherin-β Family of Nucleocytoplasmic Transport Factors; Ancient Origins and Continued Specialization. PLoS One *6*, e19308.

Odgerel, Z., Hernandez, N., Park, J., Ottman, R., Louis, E.D., and Clark, L.N. (2018). Whole Genome Sequencing and Rare Variant Analysis in Essential Tremor Families. BioRxiv 248443.

Ogawa, O., Lee, H., Zhu, X., Raina, A., Harris, P.L.R., Castellani, R.J., Perry, G., and Smith, M.A. (2003). Increased p27, an essential component of cell cycle control, in Alzheimer's disease. Aging Cell *2*, 105–110.

Oostdyk, L.T., Shank, L., Jividen, K., Dworak, N., Sherman, N.E., and Paschal, B.M. (2019). Towards improving proximity labeling by the biotin ligase BirA. Methods *157*, 66–79.

Opitz, N., Schmitt, K., Hofer-Pretz, V., Neumann, B., Krebber, H., Braus, G.H., and Valerius, O. (2017). Capturing the Asc1p/Receptor for Activated C Kinase 1 (RACK1) Microenvironment at the Head Region of the 40S Ribosome with Quantitative BioID in Yeast. Mol. Cell. Proteomics *16*, 2199–2218.

Ossareh-Nazari, B., Bachelerie, F., and Dargemont, C. (1997). Evidence for a role of CRM1 in signal-mediated nuclear protein export. Science *278*, 141–144.

Otsuka, A., and Abelson, J. (1978). The regulatory region of the biotin operon in Escherichia coli. Nature *276*, 689–694.

Paciorkowski, A.R., Weisenberg, J., Kelley, J.B., Spencer, A., Tuttle, E., Ghoneim, D., Thio, L.L., Christian, S.L., Dobyns, W.B., and Paschal, B.M. (2014). Autosomal recessive mutations in nuclear transport factor KPNA7 are associated with infantile spasms and cerebellar malformation. Eur. J. Hum. Genet. *22*, 587–593.

Panté, N., and Kann, M. (2002). Nuclear Pore Complex Is Able to Transport Macromolecules with Diameters of ~39 nm. Mol. Biol. Cell *13*, 425–434.

Park, K.-E., Inerowicz, H.D., Wang, X., Li, Y., Koser, S., and Cabot, R.A. (2012). Identification of karyopherin α1 and α7 interacting proteins in porcine tissue. PLoS One *7*, e38990.

Paschal, B.M., and Gerace, L. (1995). Identification of NTF2, a cytosolic factor for nuclear import that interacts with nuclear pore complex protein p62. J. Cell Biol. *129*, 925–937.

Patel, S.S., and Rexach, M.F. (2008). Discovering novel interactions at the nuclear pore complex using bead halo: a rapid method for detecting molecular interactions of high and low affinity at equilibrium. Mol. Cell. Proteomics *7*, 121–131.

Patel, V.P., and Chu, C.T. (2011). Nuclear transport, oxidative stress, and

neurodegeneration. Int. J. Clin. Exp. Pathol. *4*, 215–229.

Peifer, M., Berg, S., and Reynolds, A.B. (1994). A repeating amino acid motif shared by proteins with diverse cellular roles. Cell *76*, 789–791.

Pemberton, L.F., and Paschal, B.M. (2005). Mechanisms of receptor-mediated nuclear import and nuclear export. Traffic *6*, 187–198.

Perlson, E., Hanz, S., Ben-Yaakov, K., Segal-Ruder, Y., Seger, R., and Fainzilber, M. (2005). Vimentin-Dependent Spatial Translocation of an Activated MAP Kinase in Injured Nerve. Neuron *45*, 715–726.

Piñol-Roma, S., and Dreyfuss, G. (1992). Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. Nature *355*, 730–732.

Piñol-Roma, S., Choi, Y.D., Matunis, M.J., and Dreyfuss, G. (1988). Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA-binding proteins. Genes Dev. *2*, 215–227.

Pumroy, R.A., and Cingolani, G. (2015). Diversification of importin-α isoforms in cellular trafficking and disease states. Biochem. J. *466*, 13–28.

Pumroy, R.A., Ke, S., Hart, D.J., Zachariae, U., and Cingolani, G. (2015). Molecular Determinants for Nuclear Import of Influenza A PB2 by Importin α Isoforms 3 and 7. Structure *23*, 374–384.

Pyhtila, B., and Rexach, M. (2003). A gradient of affinity for the karyopherin Kap95p along the yeast nuclear pore complex. J. Biol. Chem. *278*, 42699–42709.

Rees, J.S., Li, X.-W., Perrett, S., Lilley, K.S., and Jackson, A.P. (2015). Protein Neighbors and Proximity Proteomics. Mol. Cell. Proteomics *14*, 2848–2856.

Renda, M., Baglivo, I., Burgess-Beusse, B., Esposito, S., Fattorusso, R., Felsenfeld, G., and Pedone, P. V (2007). Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. J. Biol. Chem. *282*, 33336–33345.

Rexach, M., and Blobel, G. (1995). Protein import into nuclei: association and dissociation reactions involving transport substrate, transport factors, and nucleoporins. Cell *83*, 683–692.

Rhee, H.W., Zou, P., Udeshi, N.D., Martell, J.D., Mootha, V.K., Carr, S.A., and Ting, A.Y. (2013). Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. Science (80-.). *339*, 1328–1331.

Ribbeck, K., Lipowsky, G., Kent, H.M., Stewart, M., and Görlich, D. (1998). NTF2 mediates nuclear import of Ran. EMBO J. *17*, 6587–6598.

Robbins, J., Dilworth, S.M., Laskey, R.A., and Dingwall, C. (1991). Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: Identification of a class of bipartite nuclear targeting sequence. Cell *64*, 615–623.

Rossoll, W., Jablonka, S., Andreassi, C., Kröning, A.-K., Karle, K., Monani, U.R., and

Sendtner, M. (2003). Smn, the spinal muscular atrophy-determining gene product, modulates axon growth and localization of beta-actin mRNA in growth cones of motoneurons. J. Cell Biol. *163*, 801–812.

Rother, F., Shmidt, T., Popova, E., Krivokharchenko, A., Hügel, S., Vilianovich, L., Ridders, M., Tenner, K., Alenina, N., Köhler, M., et al. (2011). Importin α7 is essential for zygotic genome activation and early mouse development. PLoS One *6*, e18310.

Roussos, P., Katsel, P., Davis, K.L., Giakoumaki, S.G., Lencz, T., Malhotra, A.K., Siever, L.J., Bitsios, P., and Haroutunian, V. (2013). Convergent findings for abnormalities of the NF-κB signaling pathway in schizophrenia. Neuropsychopharmacology *38*, 533–539.

Roux, K.J., Kim, D.I., Raida, M., and Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. J. Cell Biol. *196*, 801–810.

Roux, K.J., Kim, D.I., Burke, B., and May, D.G. (2013). BioID: a screen for protein-protein interactions. Curr. Protoc. Protein Sci. *74*, Unit 19.23.

Sakai, M., Sohda, M., Miyazaki, T., Suzuki, S., Sano, A., Tanaka, N., Inose, T., Nakajima, M., Kato, H., and Kuwano, H. (2010). Significance of karyopherin-{alpha} 2 (KPNA2) expression in esophageal squamous cell carcinoma. Anticancer Res. *30*, 851–856.

Sams, D.S., Nardone, S., Getselter, D., Raz, D., Tal, M., Rayi, P.R., Kaphzan, H., Hakim, O., and Elliott, E. (2016). Neuronal CTCF Is Necessary for Basal and Experience-Dependent Gene Regulation, Memory Formation, and Genomic Structure of BDNF and Arc. Cell Rep. *17*, 2418–2430.

Schatz, P.J. (1993). Use of Peptide Libraries to Map the Substrate Specificity of a Peptide-Modifying Enzyme: A 13 Residue Consensus Peptide Specifies Biotinylation in Escherichia coli. Bio/Technology *11*, 1138–1143.

Schmied, R., and Ambron, R.T. (1997). A nuclear localization signal targets proteins to the retrograde transport system, thereby evading uptake into organelles in aplysia axons. J. Neurobiol. *33*, 151–160.

Schrodinger LLC (2015). The PyMOL Molecular Graphics System, Version 1.8.

Shafique, M., Ismail-Khan, R., Extermann, M., Sullivan, D., Goodridge, D., Boulware, D., Hogue, D., Soliman, H., Khong, H., and Han, H.S. (2019). A Phase II Trial of Selinexor (KPT-330) for Metastatic Triple-Negative Breast Cancer. Oncologist theoncologist.2019-0231.

Sheffield, P., Garrard, S., and Derewenda, Z. (1999). Overcoming Expression and Purification Problems of RhoGDI Using a Family of "Parallel" Expression Vectors. Protein Expr. Purif. *15*, 34–39.

Shmidt, T., Hampich, F., Ridders, M., Schultrich, S., Hans, V.H., Tenner, K., Vilianovich, L., Qadri, F., Alenina, N., Hartmann, E., et al. (2007). Normal brain development in importin-α5 deficient-mice. Nat. Cell Biol. *9*, 1337–1338.

Shumaker, D.K., Solimando, L., Sengupta, K., Shimi, T., Adam, S.A., Grunwald, A., Strelkov, S. V, Aebi, U., Cardoso, M.C., and Goldman, R.D. (2008). The highly conserved nuclear lamin Ig-fold binds to PCNA: its role in DNA replication. J. Cell Biol. *181*, 269–280.

Sloan, C.A., Chan, E.T., Davidson, J.M., Malladi, V.S., Strattan, J.S., Hitz, B.C., Gabdank, I., Narayanan, A.K., Ho, M., Lee, B.T., et al. (2016). ENCODE data at the ENCODE portal. Nucleic Acids Res. *44*, D726–D732.

Smith, A., Brownawell, A., and Macara, I.G. (1998). Nuclear import of Ran is mediated by the transport factor NTF2. Curr. Biol. *8*, 1403-S1.

Stade, K., Ford, C.S., Guthrie, C., and Weis, K. (1997). Exportin 1 (Crm1p) Is an Essential Nuclear Export Factor. Cell *90*, 1041–1050.

Ström, A.-C., and Weis, K. (2001). Importin-beta-like nuclear transport receptors. Genome Biol. *2*, reviews3008.1.

Sun, B.K., Boxer, L.D., Ransohoff, J.D., Siprashvili, Z., Qu, K., Lopez-Pajares, V., Hollmig, S.T., and Khavari, P.A. (2015). CALML5 is a ZNF750- and TINCR-induced protein that binds stratifin to regulate epidermal differentiation. Genes Dev. *29*, 2225–2230.

Sun, C., Yang, W., Tu, L.-C., and Musser, S.M. (2008). Single-molecule measurements of importin alpha/cargo complex dissociation at the nuclear pore. Proc. Natl. Acad. Sci. U. S. A. *105*, 8613–8618.

Tarendeau, F., Crepin, T., Guilligay, D., Ruigrok, R.W.H., Cusack, S., and Hart, D.J. (2008). Host Determinant Residue Lysine 627 Lies on the Surface of a Discrete, Folded Domain of Influenza Virus Polymerase PB2 Subunit. PLoS Pathog. *4*, e1000136.

Tejomurtula, J., Lee, K.-B., Tripurani, S.K., Smith, G.W., and Yao, J. (2009). Role of Importin Alpha8, a New Member of the Importin Alpha Family of Nuclear Transport Proteins, in Early Embryonic Development in Cattle. Biol. Reprod. *81*, 333–342.

Theiss, A.L., Jenkins, A.K., Okoro, N.I., Klapproth, J.-M.A., Merlin, D., and Sitaraman, S. V. (2009). Prohibitin Inhibits Tumor Necrosis Factor alpha-induced Nuclear Factor-kappa B Nuclear Translocation via the Novel Mechanism of Decreasing Importin 3 Expression. Mol. Biol. Cell *20*, 4412–4423.

Thompson, E.M., Legouy, E., and Renard, J.-P. (1998). Mouse embryos do not wait for the MBT: Chromatin and RNA polymerase remodeling in genome activation at the onset of development. Dev. Genet. *22*, 31–42.

Thompson, K.R., Otis, K.O., Chen, D.Y., Zhao, Y., O'Dell, T.J., and Martin, K.C. (2004). Synapse to Nucleus Signaling during Long-Term Synaptic Plasticity: a Role for the Classical Active Nuclear Import Pathway. Neuron *44*, 997–1009.

Tran, E.J., and Wente, S.R. (2006). Dynamic Nuclear Pore Complexes: Life on the Edge. Cell *125*, 1041–1053.

Tsuji, L., Takumi, T., Imamoto, N., and Yoneda, Y. (1997). Identification of novel

homologues of mouse importin α, the α subunit of the nuclear pore-targeting complex, and their tissue-specific expression. FEBS Lett. *416*, 30–34.

Vetter, I.R., Arndt, A., Kutay, U., Görlich, D., and Wittinghofer, A. (1999). Structural View of the Ran–Importin β Interaction at 2.3 Å Resolution. Cell *97*, 635–646.

Virant-Klun, I., Leicht, S., Hughes, C., and Krijgsveld, J. (2016). Identification of Maturation-Specific Proteins by Single-Cell Proteomics of Human Oocytes. Mol. Cell. Proteomics *15*, 2616–2627.

Vuorinen, E.M., Rajala, N.K., Rauhala, H.E., Nurminen, A.T., Hytönen, V.P., and Kallioniemi, A. (2017). Search for KPNA7 cargo proteins in human cells reveals MVP and ZNF414 as novel regulators of cancer cell growth. Biochim. Biophys. Acta - Mol. Basis Dis. *1863*, 211–219.

Vuorinen, E.M., Rajala, N.K., Ihalainen, T.O., and Kallioniemi, A. (2018). Depletion of nuclear import protein karyopherin alpha 7 (KPNA7) induces mitotic defects and deformation of nuclei in cancer cells. BMC Cancer 2018 181 *18*, 325.

Wang, C.-I.C.-W.C.-L., Wang, C.-I.C.-W.C.-L., Wang, C.-I.C.-W.C.-L., Chen, C.-D., Wu, C.-C., Liang, Y., Tsai, Y.-H., Chang, Y.-S., Yu, J.-S., and Yu, C.-J. (2011). Importin subunit alpha-2 is identified as a potential biomarker for non-small cell lung cancer by integration of the cancer cell secretome and tissue transcriptome. Int. J. Cancer *128*, 2364–2372.

Wang, L., Ma, H., Fu, L., and Yao, J. (2014). Kpna7 interacts with egg-specific nuclear factors in the rainbow trout (Oncorhynchus mykiss). Mol. Reprod. Dev. *81*, 1136–1145.

Wang, P., Zhao, Y., Liu, K., Liu, X., Liang, J., Zhou, H., Wang, Z., Zhou, Z., and Xu, N. (2019). Wip1 cooperates with KPNA2 to modulate the cell proliferation and migration of colorectal cancer via a p53-dependent manner. J. Cell. Biochem. jcb.28840.

Wang, X., Park, K.-E., Koser, S., Liu, S., Magnani, L., and Cabot, R.A. (2012). KPNA7, an oocyte- and embryo-specific karyopherin α subtype, is required for porcine embryo development. Reprod. Fertil. Dev. *24*, 382–391.

Wang, Z., Lindquist, S., Tran, K., and Parker, C.S. (1998). Developmentally regulated nuclear transport of transcription factors in Drosophila embryos enable the heat shock response. Development *125*, 4841–4850.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. *46*, W296–W303.

van der Watt, P.J., Maske, C.P., Hendricks, D.T., Parker, M.I., Denny, L., Govender, D., Birrer, M.J., and Leaner, V.D. (2009). The Karyopherin proteins, Crm1 and Karyopherin β1, are overexpressed in cervical cancer and are critical for cancer cell survival and proliferation. Int. J. Cancer *124*, 1829–1840.

Weaver, L.H., Kwon, K., Beckett, D., and Matthews, B.W. (2001). Corepressor-induced organization and assembly of the biotin repressor: a model for allosteric activation of a transcriptional regulator. Proc. Natl. Acad. Sci. U. S. A. *98*, 6045–6050.

Wei, J., and Hemmings, G.P. (2005). The KPNA3 gene may be a susceptibility candidate for schizophrenia. Neurosci. Res. *52*, 342–346.

Weis, K. (2003). Regulating Access to the Genome: Nucleocytoplasmic Transport throughout the Cell Cycle. Cell *112*, 441–451.

Winnepenninckx, V., Lazar, V., Michiels, S., Dessen, P., Stas, M., Alonso, S.R., Avril, M.-F., Ortiz Romero, P.L., Robert, T., Balacescu, O., et al. (2006). Gene Expression Profiling of Primary Cutaneous Melanoma and Clinical Outcome. JNCI J. Natl. Cancer Inst. *98*, 472–482.

Xiang, S., Wang, Z., Ye, Y., Zhang, F., Li, H., Yang, Y., Miao, H., Liang, H., Zhang, Y., Jiang, L., et al. (2019). E2F1 and E2F7 differentially regulate KPNA2 to promote the development of gallbladder cancer. Oncogene *38*, 1269–1281.

Xu, Y., and Beckett, D. (1994). Kinetics of Biotinyl-5'-adenylate Synthesis Catalyzed by the Escherichia coli Repressor of Biotin Biosynthesis and the Stability of the Enzyme-Product Complex. Biochemistry *33*, 7354–7360.

Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J., et al. (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1131–1139.

Yang, Y., Muzny, D.M., Reid, J.G., Bainbridge, M.N., Willis, A., Ward, P.A., Braxton, A., Beuten, J., Xia, F., Niu, Z., et al. (2013). Clinical Whole-Exome Sequencing for the Diagnosis of Mendelian Disorders. N. Engl. J. Med. *369*, 1502–1511.

Yasuda, Y., Miyamoto, Y., Yamashiro, T., Asally, M., Masui, A., Wong, C., Loveland, K.L., and Yoneda, Y. (2012). Nuclear retention of importin α coordinates cell fate through changes in gene expression. EMBO J. *31*, 83–94.

Yasuhara, N., Shibazaki, N., Tanaka, S., Nagai, M., Kamikawa, Y., Oe, S., Asally, M., Kamachi, Y., Kondoh, H., and Yoneda, Y. (2006). Triggering neural differentiation of ES cells by subtype switching of importin-α. Nat. Cell Biol. *9*, 72–79.

Yasuhara, N., Oka, M., and Yoneda, Y. (2009). The role of the nuclear transport system in cell differentiation. Semin. Cell Dev. Biol. *20*, 590–599.

Yasuhara, N., Yamagishi, R., Arai, Y., Mehmood, R., Kimoto, C., Fujita, T., Touma, K., Kaneko, A., Kamikawa, Y., Moriyama, T., et al. (2013). Importin alpha subtypes determine differential transcription factor localization in embryonic stem cells maintenance. Dev. Cell *26*, 123–135.

Yoshitake, K., Tanaka, S., Mogushi, K., Aihara, A., Murakata, A., Matsumura, S., Mitsunori, Y., Yasen, M., Ban, D., Noguchi, N., et al. (2011). Importin-α1 as a Novel Prognostic Target for Hepatocellular Carcinoma. Ann. Surg. Oncol. *18*, 2093–2103.

Young, J.C., Major, A.T., Miyamoto, Y., Loveland, K.L., and Jans, D.A. (2011). Distinct effects of importin α2 and α4 on Oct3/4 localization and expression in mouse embryonic stem cells. FASEB J. *25*, 3958–3965.

Zhang, H., Ju, G., Wei, J., Hu, Y., Liu, L., Xu, Q., Chen, Y., Sun, Z., Liu, S., Yu, Y., et

al. (2006). A combined effect of the KPNA3 and KPNB3 genes on susceptibility to schizophrenia. Neurosci. Lett. *402*, 173–175.

Zheng, M., Tang, L., Huang, L., Ding, H., Liao, W.-T., Zeng, M.-S., and Wang, H.-Y. (2010). Overexpression of karyopherin-2 in epithelial ovarian cancer and correlation with poor prognosis. Obstet. Gynecol. *116*, 884–891.

Zhou, Z., Rawnsley, D.R., Goddard, L.M., Pan, W., Cao, X.-J., Jakus, Z., Zheng, H., Yang, J., Arthur, J.S.C., Whitehead, K.J., et al. (2015). The Cerebral Cavernous Malformation Pathway Controls Cardiac Development via Regulation of Endocardial MEKK3 Signaling and KLF Expression. Dev. Cell *32*, 168–180.